

# **On the test score gap between Roma and non-Roma students in Hungary and its potential causes**

Gábor Kertesi\*

(Institute of Economics of RCERS-HAS)

kertesi@econ.core.hu

Gábor Kézdi

(Central European University and Institute of Economics of RCERS-HAS)

kezdig@ceu.hu

## **Abstract**

Using unique data from Hungary, we assess the gap in standardized test scores between Roma and non-Roma students and show that this gap is comparable to the size of the Black-White test score gap in the United States in the 1980s. The ethnic test score gap in Hungary is nearly entirely explained by social differences in income, wealth and parental education, while ethnic factors do not play an important role. Using reduced-form regressions, we identify two major mediating mechanisms: first, the home environment of Roma children is less favorable for their cognitive development; second, Roma children face a lower quality educational environment. Comparing children with similar home environments from the same school and class, we find that the ethnic gap in test scores is insignificant. Ethnic differences in the home environment are explained by social differences, and ethnicity seems to play no additional role. While their disadvantage in accessing high-quality education is also strongly related to social differences, Roma students seem to face additional disadvantages as subjects of ethnic segregation. The results suggest that in addition to policies designed to alleviate poverty, well-designed interventions influencing these mechanisms can also improve the skill development of Roma and other disadvantaged children.

2014 January

\* Support from the CEU Institute for Advanced Study is gratefully acknowledged.

The Roma (also known as the Romani people or Gypsies) constitute one of the largest and poorest ethnic minorities in Europe. Nearly 80 percent of Roma live in former communist countries in East Central Europe. A recent study (FRA-UNDP, 2012) indicates that this population faces widespread poverty and multiple disadvantages. The employment rate among Roma aged 20 to 64 years is approximately 30 percent in most East Central European countries (FRA-UNDP, 2012). Using multiple datasets in Hungary, Kertesi and Kézdi (2010) decompose the employment gap between Roma and non-Roma in Hungary and find that the employment gap is largely explained by educational differences. Although no direct evidence is currently available on the role of skills in the employment gap between ethnic groups, skills likely play a significant role in ethnic employment gap between ethnic groups in East Central Europe. Understanding the extent and the origins of the gap in skills between ethnic groups is therefore important for understanding the origins of the disadvantages faced by the Roma minority and developing effective policies to address such disadvantages.

This study quantifies the test score gap between Roma and non-Roma students in Hungary and aims to explain this test score gap through policy-relevant factors. We focus on two major questions: Does the test score gap between Roma and non-Roma students result from ethnic specificities of the Roma or social disadvantages? Moreover, what are the mechanisms behind the emergence of the test score gap? A brief analysis of these questions was published in an earlier study of ours (Kertesi and Kézdi, 2011). In this paper, we place the problem in a wider context, examine the mediating mechanisms in detail and form appropriate policy conclusions.

The existence of a Roma – non-Roma school achievement gap frequently leads researchers to seek an explanation related to characteristic ethnic behavior patterns. If this diagnosis were correct, the appropriate policy response should target such characteristic ethnic behavior patterns to "shape attitudes" and transform the "Roma mentality." If the achievement gap can be almost entirely explained by well-defined social differences, however, interventions intended to transform the "characteristic mentality" are likely to be ineffective. Answering the second question is equally important: finding that a disadvantaged family background is responsible for skill deficits does not provide a complete explanation. Without understanding the mediating mechanisms between poverty and low achievement, we cannot design effective policies.

This paper shows that the gap in standardized test scores in Hungary is substantial (similar to the Black-White gap in the United States in the early 1980s) and is in large part explained by social differences in income, wealth and parental education but that ethnic factors do not play an important role. We examine three mechanisms in detail and find that two of them are primarily responsible for the achievement gap between Roma and non-Roma students. Differences in health seem to play a limited role in this achievement gap, but differences in the home environment and school quality appear to be important. The home environment and parenting practices can explain, according to our regression results, one-third to two-thirds of the test score gap. We also show that the gap between Roma and non-Roma students attending the same school in the same classroom is 60 percent smaller than the national gap. When comparing children with similar home environments from the same school and class, we find that the ethnic gap in test scores becomes insignificant. Ethnic differences in the home environment are completely explained by social differences, and ethnicity in itself seems to play no additional role. However, while access to higher quality schools is strongly related to social differences, Roma students, as subjects of ethnic segregation, seem to

face additional educational disadvantages.

## **Data**

Standardized competence test scores and a survey with ethnic identifiers linked to these test score data provide a unique opportunity to analyze the test score gap between Roma and non-Roma students in Hungary. The source of these test score data is the May 2006 National Assessment of Basic Competences (NABC), which is administered to every 8<sup>th</sup> grade elementary school student. These administrative data cover the entire population of 8<sup>th</sup> grade students but contain no ethnic markers. Ethnicity, together with a wealth of family background data, is measured in a survey linked to those test scores, the Hungarian Life Course Survey (HLCS) of the Tárki Research Institute of Hungary.

The HLCS is a panel survey that follows 10,000 youths on an annual basis, beginning in the fall of 2006. The survey sampled regular students who participated in the NABC and special needs students who did not participate in the NABC but who completed a simplified version of the reading comprehension test. Students with lower test scores and special needs students are overrepresented in the sample, and we use sampling weights throughout the analysis to restore national representativeness.

The questions in the first wave of the HLCS in 2006 focused on the respondents' family structure, financial situation, early childhood experiences, medical and school history and plans for secondary school. Subsequent waves of the survey primarily concentrated on school careers and the mechanisms underlying student dropout.

In this paper, we consider data collected in the first two waves of the survey. We restricted the sample to individuals who participated in both waves of the survey and who were living with at least one of their biological parents. These sample restrictions were necessitated by the methodology that we employed to identify Roma ethnicity. The parents were asked what nationality or ethnicity they identified with primarily or secondarily in both waves of the survey. These two questions allowed the participants to choose a dual identity. For the purposes of this study, we consider a young person to be Roma if he or she had at least one biological parent who identified primarily or secondarily as Roma in either the 2006 or 2007 survey. Using this definition, Roma youth comprise nearly 8 percent of all 8<sup>th</sup> grade students; the size of the Roma subsample is 848 students (the fractions are weighted by sampling weights; see Table A1 of the Appendix). The total sample size is 9056 students with reading comprehension test results and 8335 students with mathematics test results. The difference in samples occurred because special needs students only completed the reading comprehension test.<sup>1</sup> Table A2 of the Appendix reports the magnitude of the bias arising from sample selection and the basic data on the students who were eliminated from the sample for various reasons.

## **The test score gap between Roma and non-Roma students in Hungary**

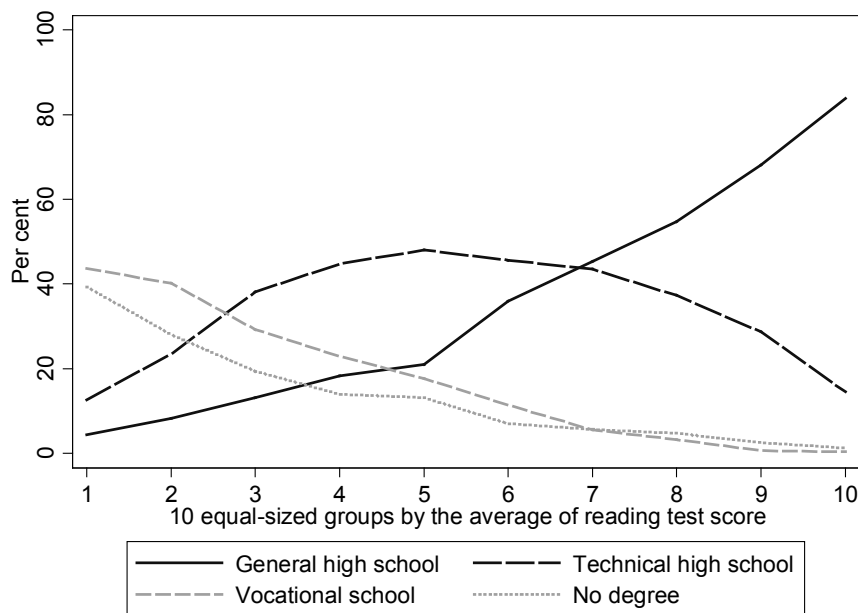
---

<sup>1</sup> Of all 8<sup>th</sup> graders, 6 percent (and 12 percent of Roma 8<sup>th</sup> graders) were special needs students in 2006; the majority were classified as having a mild intellectual disability.

As *Figure 1* shows, the test scores measure skills that have a substantial impact on the choice of secondary school and key events in the secondary school career. The figure depicts the probabilities of completing different types of secondary school by age 21 as a function of 8<sup>th</sup> grade test scores, using data from the sixth wave of the HLCS. The vertical axis indicates the fraction of respondents with a general high school degree, technical high school degree (these two degrees involve passing a graduation examination<sup>2</sup> that is also the entry test for college) or vocational school degree, as well as the fraction of respondents without any secondary degree. The horizontal axis depicts 10 equally sized categories created by the reading test scores measured in 8<sup>th</sup> grade, such that group 1 has the lowest and group 10 the highest scores.

**Figure 1**

**The likelihood of acquiring different types of secondary school degrees by age 21 as a function of 8<sup>th</sup> grade reading comprehension test scores**



*Figure 1* reveals that a strong, skill-based selection mechanism is at work in secondary school. The relationship between the likelihood of lacking a secondary degree and test scores is monotonically negative, and the relationship between the likelihood of earning a vocational degree and test scores is very similar. The likelihood of earning a general high school degree is strongly positively related to test scores. The results presented in *Figure 1* imply that selection into secondary school types and subsequent success is strongly related to skills in 8<sup>th</sup> grade and that the NABC test scores are good measures of those skills. Moreover, labor market prospects are strongly related to the type of secondary school degree. Between 2006 and 2012, the employment rate at age 30 was 50 percent for those with 8 grades of education only, compared to over 75 percent for those with some type of secondary degree. The wages of vocational school graduates were 20 percent higher, the wages of technical high school graduates were 180 percent higher, and the wages of college graduates (the degree obtained by most of the general high school graduates) were over 200 percent higher than the wages of those with only 8 grades of education. Taken together, these results provide strong

<sup>2</sup> Called „maturity exam” in Hungary. Comparable to the A-level exams in the United Kingdom.

evidence for the importance of the test score differences.

We now turn to ethnic differences in the test scores. *Table 1* reports the magnitude of the standardized test score gap between Roma and non-Roma students using the data on 8<sup>th</sup> graders from the 2006 NABC. As a comparison, we provide similar data on the test score gap between 13-year-old and 8<sup>th</sup> grade Black and White students in the United States. We include the test scores of 13-year-old students from the United States because this is the format of the data from the recent past, at the turn of the 1980s.

**Table 1**  
**The magnitude of the Roma–non-Roma test score gap in Hungary and the black-white test score gap in the United States (measured in standard deviations of the national average of the given test)**

Year	Roma–non-Roma gap, 8 <sup>th</sup> grade, Hungary <sup>a</sup>		Black-White gap, 8 <sup>th</sup> grade, United States <sup>b</sup>		Black-White gap, 13-year-olds, United States <sup>c</sup>	
	reading	mathematics	reading	mathematics	reading	mathematics
1978/80	–	–	–	–	–0.91	–1.08
1992	–	–	–0.83	–1.10	–0.73	–0.93
2006/8	–0.97	–1.05	–0.78	–0.88	–0.56	–0.81

<sup>a</sup> Calculated by the authors. *Source:* the combined data of the 2006 NABC and the HCLS.

<sup>b</sup> National Assessment of Educational Progress (NAEP), Main NAEP tables, 1992 and 2007.

<sup>c</sup> NAEP, Long-Term Trend tables, reading: 1980, 1992 and 2008, mathematics: 1978, 1992 and 2008.

The difference between Roma and non-Roma students' scores is approximately one standard deviation. This difference matches the size of the gap between Black and White 13-year-old students in the United States in about 1980, which has narrowed significantly since.

While our data only cover 8<sup>th</sup> graders, we can shed some light on the age pattern of the gap with the use of other, albeit not nationally representative, data (*Table 2*). The coverage of the samples and, in one case, the tests differ from those in our data. As a result, meaningful comparisons across age groups can only be made within each sample.

The first data come from the evaluation of the National Education Integration Network program (Kézdi and Surányi, 2008). These data enable us to compare 2<sup>nd</sup> and 4<sup>th</sup> grade students. The study measured the arithmetic and reading skills of approximately 4000 students in 60 schools in two waves (spring 2005 and spring 2007). The tests were developed for the study, and disadvantaged students are highly overrepresented in the sample. The second data allow us to compare 6<sup>th</sup> and 8<sup>th</sup> grade students; these data are based on the "Interethnic Relations, 2010" survey. The survey collected data on 8<sup>th</sup> grade students at 88 schools, and the respondents were linked to their administrative files with their 6<sup>th</sup> grade test scores from 2008 and the 8<sup>th</sup> grade test scores from 2010. The tests are the standard NABC tests, and again, disadvantaged students are overrepresented in the sample. The third dataset allows us to compare 8<sup>th</sup> and 10<sup>th</sup> graders: sample is the subsample of the HLCS that was matched to the 10<sup>th</sup> grade test scores of the NABC data. Owing to imperfections in the matching procedure, this subsample is 50 percent of the original sample, and students of higher status are slightly overrepresented in the sample.

**Table 2**

**The Roma–non-Roma test score gap by grade level (measured in standard deviations of the national average of the given test). First dataset: raw gap; data in parentheses: include controls for gender, age, no mother/father and parental education**

Survey/Year	Grade	Test		
		SZTE arithmetic <sup>d</sup>	SZTE reading <sup>d</sup>	NABC reading <sup>e</sup>
OOIH <sup>a</sup>	2 <sup>nd</sup>	-0.76 (-0.49)	-	-
2005/2007	4 <sup>th</sup>	-	-0.86 (-0.53)	-
IEK-OKM <sup>b</sup>	6 <sup>th</sup>	-	-	-0.67 (-0.33)
2008/2010	8 <sup>th</sup>	-	-	-0.68 (-0.35)
HLCS-OKM <sup>c</sup>	8 <sup>th</sup>	-	-	-0.82 (-0.22)
2006/2008	10 <sup>th</sup>	-	-	-1.01 (-0.33)

<sup>a</sup> The evaluation of the National Education Integration Network (OOIH) program; sample: students in 2<sup>nd</sup> grade in spring 2005 and 4<sup>th</sup> grade in spring 2007. See: Kézdi and Surányi, 2008.

<sup>b</sup> The sample of the "Inter-Ethnic Relations, 2010" (IEK) in Education survey combined with the 2008 NABC 6<sup>th</sup> grade and the 2010 NABC 8<sup>th</sup> grade test score data.

<sup>c</sup> The sample of the Tárki HLCS is combined with the 2006 NABC 8<sup>th</sup> grade and 2008 NABC 10<sup>th</sup> grade test score data. The table only includes data on students from the HLCS if they could be identified as 10<sup>th</sup> graders in the 2008 NABC.

<sup>d</sup> Reading comprehension test for 2<sup>nd</sup> graders and arithmetic skills test for 4<sup>th</sup> graders developed by the Institute of Education at the University of Szeged. The national mean and standard deviation data are from the longitudinal survey of the Institute of Education, University of Szeged, sample III, 2005: 2<sup>nd</sup> graders, 2006: 4<sup>th</sup> graders. (See: Csapó, 2007)

<sup>e</sup> NABC reading comprehension and mathematics tests.

We summarize the results of all measurements in *Table 2*. In addition to the raw test score gap, we include the values of the gap after we corrected for gender, age, household presence and education of the mother/father in parentheses.

The available data indicate the relative stability of the test score gaps measured in grades 5 to 8, but the gap increases between grades 2 and 4 and grades 8 and 10. As the gap in the reading test scores is generally larger, the observation that the reading gap is larger in 4<sup>th</sup> grade than the math gap in 2<sup>nd</sup> grade suggests an even larger increase in the gap concerning the scores on each test. Conditioning on parental education leads to substantially smaller gaps, especially in higher grades, and these conditional gaps appear to widen, too.

International surveys (Lee and Burkam, 2002; Neuman, 2006) find that the children of disadvantaged minorities struggle with significant deficits by the time that they reach kindergarten age. The available evidence is scarce but suggests that poor children in Hungary are no exceptions to this rule. The evaluation of the early childhood education program Biztos Kezdet (Sure Start) in Hungary collected baseline data on 4- to 6-year-old kindergarteners and measured the vocabularies of these children. In this sample, the raw gap between Roma and non-Roma children is 66 percent of a standard deviation, which is reduced to 11 percent once we condition on gender, age, household presence and education of the mother/father.

### **Social composition and the achievement gap**

What is the magnitude of the ethnic gap compared to the raw test score gap if we account for social and income differences between the Roma and non-Roma student populations? As non-Roma students constitute a much larger percentage of the students (and thus, of the sample), we conduct the following thought experiment: how large would the test score gap between Roma and non-

Roma students be if non-Roma students lived in similarly poor conditions to those in which Roma students live?

In our analysis, we used the family background variables presented in Table 3. In conjunction, these variables represent the family's long-term income, wealth and life chances in a broad sense. The descriptive statistics are presented in Table A3 of the Appendix.

**Table 3**  
**Family background variables**

Variable name	Definition of variable
Biological mother in household	Lives with biological mother: yes/no
Nonbiological mother in household	Lives with nonbiological mother: yes/no
Biological father in household	Lives with biological father: yes/no
Nonbiological father in household	Lives with nonbiological father: yes/no
Mother's education	Mother's (biological/nonbiological) highest completed level of education: 0-8 years of elementary school/vocational school/high school diploma/higher education
Father's education	Father's (biological/nonbiological) highest completed level of education: 0-8 years of elementary school/vocational school/high school diploma/higher education
Mother's current employment	Mother was employed in the fall of 2006: yes/no
Father's current employment	Father was employed in the fall of 2006: yes/no
Mother's long-term employment	Mother: share of years worked while the child was 0-14 years old, %
Father's long-term employment	Father: share of years worked while the child was 0-14 years old, %
ln(monthly income)	The logarithm of the household's monthly income, 2006
ln(number of household members)	The logarithm of the number of household members
Number of unemployed adults	Number of unemployed adult household members
Living space per person, m <sup>2</sup>	Surface area of apartment/number of household members, m <sup>2</sup> /person
Number of rooms per person	Number of rooms/number of household members
Bathroom	Is there a bathroom in the apartment? yes/no
Poverty1 (income does not cover food)	Was there not enough money for food in the past 12 months? yes/no
Poverty2 (income does not cover heating)	Was there not enough money for heating in the past 12 months? yes/no
Poverty3 (child-care assistance)	The family receives child-care assistance: yes/no
Poverty4 (free school meals)	The child receives free meals at school: yes/no
Poverty5 (free school textbooks)	The child receives free textbooks at school: yes/no
Place of residence: region	Regions of Hungary: Central Hungary/Central Transdanubia/Western Transdanubia/Southern Transdanubia/Northern Hungary/Northern Great Plain
Place of residence: type	Budapest/county seat/other city/village
Place of residence: remote	Access to the place of residence is too expensive or time consuming by car or public transport <sup>a</sup> : yes/no

<sup>a</sup> See *Köllő* [1997].

We summarize the results of various estimations in *Table 4*. We estimate the role of social background in the achievement gap between Roma and non-Roma students using two methods: linear regression (OLS) and propensity score matching. The OLS results are more standard, but propensity score matching is more flexible, as it allows for nonlinearities and ensures common support. We estimate two types of matching models: nearest neighbor matching and stratified matching.

**Table 4**  
**The magnitude of the ethnic test score gap conditional on social background.**  
**Regression and matching estimates**

	Roma parameter (standard error) <sup>a</sup>	Number of observations <sup>b</sup>	R <sup>2</sup>
<i>Reading comprehension</i>			
Raw gap	-0.97 (0.05)**	9056	0.06
OLS	-0.23 (0.05)**	9056	0.27
Propensity score matching			
nearest neighbor matching	-0.18 (0.06)*	837/480	–
stratified matching	-0.18 (0.04)*	837/7948	–
<i>Mathematics</i>			
Raw gap	-1.05 (0.05)**	8335	0.07
OLS	-0.32 (0.05)**	8335	0.27
Propensity score matching			
nearest neighbor matching	-0.26 (0.06)*	837/395	–
stratified matching	-0.26 (0.04)*	837/7948	–

<sup>a</sup> Standard errors in parentheses.

<sup>b</sup> In the case of propensity score matching: number of Roma (treatment)/non-Roma (control) observations

\* Significant at 5 %, \*\* Significant at 1 %.

*Note:* see detailed results in Table A4 of the Appendix.

Despite the methodological differences, all estimates show that the test score gap between Roma and non-Roma students is to a large extent explained by their adverse long-term socio-economic conditions. The test score gap between the average Roma student and the average non-Roma student is approximately one standard deviation in magnitude. The test score gap between Roma students and non-Roma students of similar social backgrounds is approximately 0.2-0.3 of a standard deviation. One way to interpret these findings is that three-fourths of the raw mathematics gap and four-fifths of the raw reading comprehension gap would disappear if Roma and non-Roma students had similar social backgrounds. Many non-Roma students have similarly disadvantaged backgrounds to those of the average Roma student; however, few Roma students have backgrounds that are similar to or better than the average non-Roma student. Our results are therefore identified among the bottom of the social background distribution of non-Roma students.

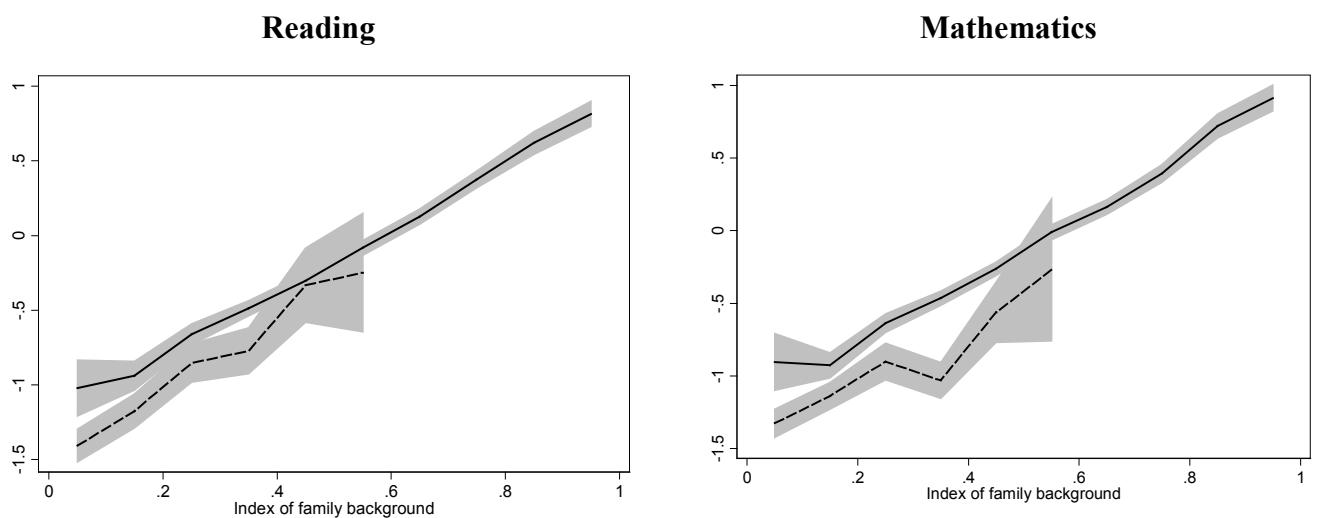
Is there a way to address whether the ethnic test score gap would decline significantly if Roma students lived in conditions that were similarly as good as those of the average or better-than-average non-Roma student? Answering this question requires an extrapolation of the test score gap as a function of social background. We created a one-dimensional synthetic family background index by taking the linear combination of all our family background variables through the use of coefficients obtained from a regression of test scores (the average of the reading and mathematics scores) on the family background variables. We then normalized the resulting values on a range



from 0 to 1. Individuals facing worse socio-economic circumstances are thus located closer to 0, while those living in better conditions are closer to 1. Figure A1 in the Appendix plots the distribution of the family background index for the Roma and non-Roma subsamples separately. The overwhelming majority of Roma students live in worse conditions than the average non-Roma student: the Roma subsample is too small to be meaningful over values of 0.6.

We divided the range of the family background index into 10 equal intervals and estimated the mean Roma and non-Roma reading and mathematics test scores for each interval. We restricted the estimates for the Roma students to the 0–0.6 range. The estimates are presented in *Figure 3*. The gray zone indicates the 95 percent confidence intervals (within  $\pm 2$  standard errors of the mean).

**Figure 3**  
**Reading and mathematics test results as a function of the family background index**  
 (The gray zone indicates the 95 percent confidence intervals)  
 Continuous lines: Non-Roma. Dashed lines: Roma



Although our method would allow for nonlinear relationships, the relationship between the family background index and expected test scores is nearly linear for both the Roma and the non-Roma samples. The two lines are also very close to one another. In the case of the reading score, the difference is very small and tends to decrease as the family background index values increase; in the case of the mathematics score, the difference is somewhat larger, and it is difficult to determine whether the two lines converge or diverge. Extrapolating beyond the common support, these results suggest that the test scores of Roma students would be similar to, or only slightly worse than, the better-off non-Roma students if their social circumstances were also similar.

We have therefore answered our first question: the test score gap between Roma and non-Roma students in Hungary are, to a large extent, explained by social background, while ethnicity seems to play a very small role, at most, in the test score gap. We now turn to answering our second question: What mediating mechanisms are responsible for the relationship between social background and test scores that lead to the large test score gap between Roma and non-Roma students?

## Theoretical framework

The literature identifies three major sets of mechanisms that lead to low achievement among disadvantaged students. In general, children's skill accumulation and school performance are weaker if (1) their health is worse than average, (2) they have little access to resources and activities that are important for developing their skills in their home environment and (3) they have limited access to high-quality educational services and a motivating school environment. We review the international evidence on these mechanisms in this section.

1. *Health.* Pain, fatigue and stress associated with poor health and diseases have a direct effect on learning performance. Missed lessons reduce the time spent studying, and parents are often overprotective of more vulnerable children, allowing them to spend less time in the company of their peers and providing them with fewer opportunities for sports and other activities that can help to develop their skills (Currie, 2005; Case, Lubotsky and Paxson, 2002; Almond and Currie 2011). Unfavorable circumstances during pregnancy/birth and chronic disease during early childhood create the conditions for diseases in later stages of childhood and adulthood and have a negative effect on the development of the skills necessary for learning (Barker, 1998; Reichman, 2005; Case, Fertig and Paxson, 2005; Palloni et al., 2009).

The children of less-educated and poor families have a higher than average risk of contracting chronic diseases and suffering accidents and injuries. Parents are also less likely to recognize the symptoms of disease, and such families have more limited access to better health care owing to insufficient information and transportation and other costs. Poor children thus have a more difficult time recovering from diseases. As a result, children of poor families are of systematically worse health on average than their higher-income counterparts, and this difference appears to increase with age (Case, Lubotsky and Paxson, 2002; Currie and Stabile, 2003; Currie 2009).

2. *Home environment/parenting.* The numerous activities, tools and aspects of the material environment and behavioral patterns combine to form the learning environment at home. We focus on two components: (1) the availability of activities, objects, tools and environmental factors that directly or indirectly promote the child's cognitive development and (2) parenting practices that guarantee the child's emotional stability (Linver, Brooks-Gunn and Kohen, 2002). The literature offers two theories to explain the relationship between these mechanisms and poverty. *Human capital theory*<sup>3</sup> asserts that a low level of parental investment is responsible for the negative impact of the parental poverty on children's skills. The effects of poverty on a child's human capital (in a broad sense) are thus mediated by tools, experiences and parental "services" that stimulate the child's development. The *family stress model*<sup>4</sup> asserts that economic hardship or the loss of a job influences children's development through the parents' mental state. As the parents' mental state affects the parent-child relationship and the parenting methods that are used in the family, it has a major impact on children's development. The two classes of explanations are, to some extent, competing theories, but they complement each other in many respects.

---

<sup>3</sup> Leibowitz, 1974; Becker, 1981a; 1981b; Becker and Tomes 1986; Haveman and Wolfe, 1995; Mayer, 1997; Mulligan, 1997; Kalil and DeLeire, 2004; Guryan, Hurst and Kearney, 2008; Gould and Simhon, 2011; Kaushal, Magnuson and Waldfogel, 2011; Phillips, 2011.

<sup>4</sup> Elder, 1974; Lempers, Clark-Lempers and Simons, 1989; McLoyd, 1990; Conger et al., 1992; 1993.

3. *School quality.* Two central factors can make a school a "high-quality" institution: effective teachers and mutually motivating classmates. Although measuring teaching quality is difficult, a number of innovative studies conducted over the past two decades have convincingly demonstrated that teacher performance plays a definitive role in students' school performance. These studies assess teaching quality through the use of a variety of methods: some measure observable features, such as the results of teacher skill tests (Ferguson, 1998), others measure student performance with value added models (Rivkin, Hanushek and Kain, 2005; Chetty, Friedman and Rockoff, 2011), and yet others compare the outcomes of up-to-date and obsolete teaching practices in the classroom (Wenglinsky, 2001; Schacter and Thum, 2004). Their results are clear: high-quality teaching is one of the main catalysts for good student performance.

Peer group composition is positively related to student performance. If any type of social mechanism causes children with learning problems to cluster in one school or classroom, a subculture may develop that is not conducive to learning. The leaders of the peer group may refuse to make an effort and co-operate with the teachers and create their own culture of resistance to school knowledge (Akerlof and Kranton, 2002; Bishop et al., 2003; Fryer and Torelli, 2010). A number of studies indicate that high-performance peer groups enhance while low-performance peer groups inhibit individual learning performance (Ammermueller and Pischke, 2009; Hanushek et al., 2003; Hanushek, Kain and Rivkin, 2009).

Teacher quality and peer group composition may be positively correlated, which may hinder the separation of their respective effects, on the one hand, but may result in mutually reinforcing effects, on the other. Such a positive correlation is more likely in school systems that are characterized by higher levels of segregation and that do not compensate teachers for more difficult tasks associated with teaching more difficult peer groups.<sup>5</sup> Recent studies show the consequences of the negative selection of teachers to worse performing schools in segregated school systems (Hanushek, Kain and Rivkin, 2004; Clotfelter, Ladd and Vigdor, 2005; Jackson, 2009).

*Figure 5* summarizes our theoretical framework regarding the factors that may contribute to the Roma students' school deficits. For the sake of simplicity, the figure treats the social background of a student as one dimensional, namely, good or bad; similarly, health, home environment and school quality are captured by one-dimensional variables that are also binary. The simple lines (not arrows) connecting the variables designate correlations, the arrows designate causal relationships, and the plus and minus symbols indicate the signs of the relationships.

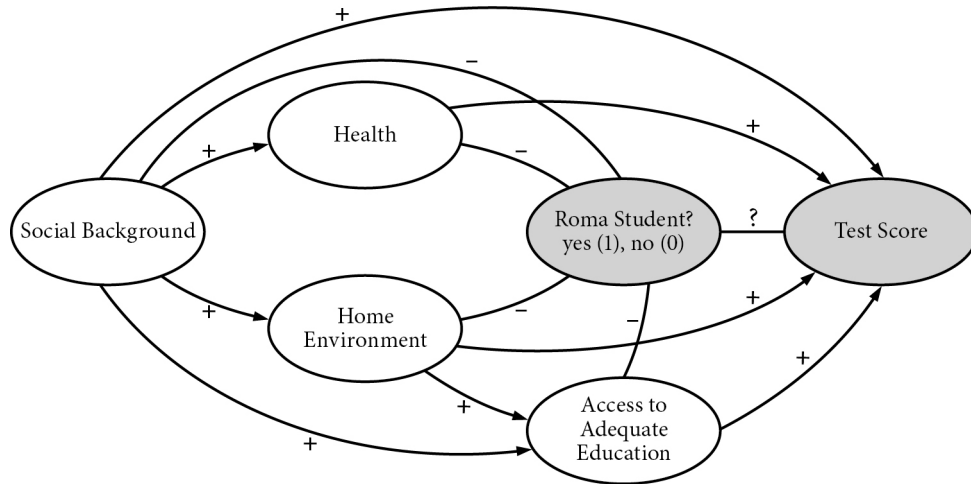
Social background is treated as a predetermined characteristic that can influence children's health, the home learning environment, access to quality education and test scores; reverse causality is unlikely to be very important in this case. The role of ethnicity is more complex. With respect to the relationship between ethnicity and social background, causality can run in both directions (e.g., social background may affect identity, and ethnic discrimination may affect life chances). Causality may also run in both directions for the relationship between ethnicity and the intervening variables representing the transmission mechanisms (e.g., school environment may affect identity, and ethnic

---

<sup>5</sup> Well-designed social experiments and additional resources can achieve good results with disadvantaged students in schools that are dominated by disadvantaged students (see, for example: Dobbie and Fryer, 2011 on the Harlem Children's Zone and Angrist et al., 2010 on the Knowledge is Power program).

segregation in schools may affect the quality of education).

**Figure 5**  
**An illustration of the causal relationships that determine test results**



### Measurement strategy

The lack of exogenous variation in health, home environment, parenting practices and the school environment prevents us from performing a causal analysis. Instead, we use the richness of our data to apply as detailed measures of each factor as possible and conduct a decomposition exercise to assess the potential magnitude of each—conditional on each other. The HLCS data provide us with the following measures.

1. *Student health* is measured by two variables: birth weight and self-reported health. Birth weight is one of the most important indicators that characterize pregnancy conditions and fetal development. Children born with a low birth weight—measured as a birth weight under 2500 g—have a higher risk of physical and nervous system damage; have a higher likelihood of developing learning difficulties, attention deficit problems and special educational needs; are more likely to repeat grades and have lower test scores (Breslau et al., 1994; Hack, Klein and Taylor, 1995; Reichman, 2005). In addition to correlations, several studies show the causal effects of low birth weight on education levels, employment chances and incomes (Currie and Hyson, 1999, Behrman and Rosenzweig, 2004; Black, Devereux and Salvanes, 2007; Oreopoulos et al., 2008). The incidence of low birth weight is closely correlated with the income, wealth and education of the population concerned. The poorer and less educated the population of a country or a group within a country is, the greater the statistical probability of low birth weight will be (Behrman and Rosenzweig, 2004) owing to various mechanisms, including nutrition, health behavior and access to health care (Rosenzweig and Schultz, 1982; Hack, Klein and Taylor, 1995; Cramer, 1995; Meara, 2001; Schonkoff and Phillips, 2004, chapter 8; Paul, 2010; Currie, 2011).

The second variable is the self-reported health of the surveyed students. It was measured on a scale from one to four (excellent/good/adequate/poor) a few months after the reading and mathematics tests were taken. This variable, which is widely used in the literature, is strongly correlated with

both medically diagnosed chronic conditions (Case, Lubotsky and Paxson, 2002) and parental social status indicators (income and education). Poorer children generally tend to have worse health, which is reflected in their self-evaluations, or, in the case of younger children, in their parents' subjective evaluations (Case, Lubotsky and Paxson, 2002; Currie and Stabile, 2003; Case, Fertig and Paxson, 2005; Currie, 2009, Table 1).

2. In assembling the *parenting/home environment* indicators, we used retrospective questions in the HLCS dating back to kindergarten. We also used a series of questions and observations in the first wave of the HLCS to measure the material and emotional home environment in adolescence.

Early childhood experiences and family interactions related to books and other written texts play an exceptionally important role in children's cognitive development. Regular bedtime storytelling sessions and parent-child interactions centered on browsing children's books together (including picture books) are important ways in which toddlers and kindergarteners acquire such experiences. The number of literacy experiences in early childhood can have an important effect on the child's basic skills prior to school enrollment (Heath, 1983; Réger, 1990; Neuman, 1996; Sénéchal et al., 2001; Dickinson and Tabors, 2001; Raikes et al., 2006). We have two measures of the frequency of bedtime storytelling sessions at kindergarten age in the HLCS, one from the parents and one—in a separate interview—from the children. The HLCS also contains questions on other joint activities, of which hiking or engaging in sports was significantly related to test scores and hence is included in our analysis.

The students' current home environment and parenting practices are measured with the use of the HOME (*Home Observation for Measurement of the Environment*) scale. The HOME index is an instrument that is used to assess the developmentally relevant features of a child's home environment, and our data contain the battery developed for adolescents (Bradley et al., 2000; Mott, 2004). Recent research shows that the home environment and parenting, as measured by the HOME scale, are strongly related to children's school readiness and subsequent school performance (Crane, 1996; Guo and Harris, 2000; Linver, Brooks-Gunn and Kohen, 2002; Brooks-Gunn and Markman, 2005; Todd and Wolpin 2007). The first wave of the HLCS, in 2006, relied on an adapted version of the short form of the adolescent HOME scale (HOME-SF) that was used in the National Longitudinal Study of Youth. The short version is composed of 27 items and assesses two subscales: cognitive stimulation and emotional support. As supplemental measures of the home environment, we also included a key variable of the PISA studies (the number of books in the home) and information on the availability of an internet connection.

We describe the variables that characterize students' health and home environment and parenting in *Table 5*. *Table A5* in the Appendix presents the summary statistics for Roma and non-Roma students for these variables.

**Table 5**  
**Intervening variables representing the transmission mechanisms**

Variable name	Definition
<b>HEALTH</b>	
Low birth weight	The child was born with a birth weight lower than 2500 g: yes/no
Adequate or poor teenage health	The child's health, based on a fall 2006 <i>self-evaluation</i> , is adequate or poor according to a four-part scale (poor/adequate/good/excellent): yes/no (modal age: 15)
<b>HOME ENVIRONMENT/PARENTING</b>	
Seldom or never told bedtime stories (child's response)	Seldom or never told bedtime stories (once every 6 months or even less frequently) while the child was in kindergarten: yes/no (child's response)
Often told bedtime stories (child's response)	Often told bedtime stories (several times a week) while the child was in kindergarten: yes/no (child's response)
Seldom or never told bedtime stories (parent's response)	Seldom or never told bedtime stories (never or almost never) while the child was in kindergarten: yes/no (parent's response)
Often told bedtime stories (parent's response)	Often told bedtime stories (every day or almost every day) while the child was in kindergarten: yes/no (parent's response)
Seldom went hiking with parents (child's response)	Seldom (once every 6 months or even less frequently) went hiking or engaged in sports together with the parents while the child was in kindergarten: yes/no (child's response)
HOME index, cognitive subscale <sup>a</sup>	The subscale of the HOME index (a synthetic variable characterizing the home environment) for 15-year-olds that measures cognitive stimulation
HOME index, emotional subscale <sup>a</sup>	The subscale of the HOME index (a synthetic variable characterizing the home environment) for 15-year-olds that measures emotional support
Number of books at home	The number of books in the home: under 50/50-150/150-300/300-600/600-1000/over 1000
Internet connection at home	Does the home have an internet connection: yes/no

<sup>a</sup> Table A6 of the Appendix presents the items in the HOME index's cognitive and emotional subscales.

3. In contrast to health and home environment, we do not use explicit measures to capture the potential effects of *school quality*. Instead, we compare Roma and non-Roma students who studied in the same school and class with the use of including school and class fixed effects. Note that in general, assignment to classes (groups of 20 to 30 students) is fixed over a student's entire school career, and hence, students in the same class generally share a common school history. Recall that our data are linked to the administrative NABC database, which contains the students' school and class identification numbers in addition to their test scores. The multistage sampling method and size of the HLCS sample yield a sufficient number of observations for within-class analysis.

When interpreting the results, we can interpret the regression estimates of the "Roma" coefficient in the equations without school and class fixed effects to measure the differences between randomly selected Roma and non-Roma students. The "Roma" coefficient in the equations that include school and class fixed effects measures the gap between randomly selected Roma and non-Roma *classmates*. The difference between the two estimates measures the test score gap between Roma and non-Roma students who are *not classmates*. This residual component incorporates the consequences of the selection of typical Roma students into schools and classes that differ from the schools and classes of typical non-Roma students. This residual component thus captures all the effects of selection and differences in the educational quality of typical Roma and non-Roma

students. The estimate is an upward-biased estimate of the effects of school quality because of selection: the non-Roma classmates of most Roma students are likely to differ from the average non-Roma student. We partially control for this bias by including the rich set of family background variables, but the remaining estimates are likely to remain larger than the true effect of school quality.

### The strength of the transmission mechanisms

Our first question concerns the relative importance of the three transmission mechanisms as the basic pathways between social background and the ethnic achievement gap. These three mechanisms are strongly interrelated, in part because of unobserved factors. As a result, a multiple regression model that includes all covariates and fixed effects can yield informative results regarding the potential combined effect of the three mechanisms but not regarding their relative magnitudes. Successive inclusion of the variables representing these mechanisms also generates concern, as the order in which the variables enter matters. Therefore, we enter the variables representing health, the home learning environment and school/class fixed effects into the equation in varying order, and finally, we enter the family background variables that characterize the family's socio-economic conditions.

**Table 6**  
**The magnitude of the residual ethnic test score gap after accounting for the transmission mechanisms**

	Reading			Mathematics		
Roma	-0.97 (0.05)**	-0.07 (0.07)	-0.05 (0.07)	-1.05 (0.05)**	-0.18 (0.07)**	-0.15 (0.07)*
Health, home environment	–	yes	yes	–	yes	yes
School/class fixed effect	–	yes	yes	–	yes	yes
Family background	–	–	yes	–	–	yes
Sample size	9056	9056	9056	8335	8335	8335
R <sup>2</sup>	0.06	0.67	0.68	0.07	0.68	0.69

Standard errors in parentheses are clustered at the school level.

\* Significant at 5 %, \*\* Significant at 1 %.

Note: see detailed results in Table A7 of the Appendix.

We first examine the combined effect of the three mechanisms. As *Table 6* shows, the bulk of the raw test score gap *disappears* (over 90 percent of the reading and over 80 percent of the mathematics test score gap) if we account for our measures of the three mechanisms. No gap in reading and a small gap in mathematics exist between Roma and non-Roma students who are similar in terms of health, who had the same degree of access to the resources, tools and activities that stimulate skill development in their home environment and who attended the same classes in the same schools. Entering the family background variables does not significantly reduce the Roma coefficient once the mechanism variables are included. If interpreted as causal effects, the results suggest that the skill deficits of Roma students are exclusively due to well-defined social mechanisms related to health, home environment and educational quality.

Our second question concerns the relative strength of the three mechanisms. *Table 7* presents our estimates for the potential of each mechanism to explain the test score gap between Roma and non-Roma students. The table reports our lower and upper estimates. The lower estimates correspond to

the reduction in the Roma coefficient in the test score regression when the variables of the particular mechanism are entered last (when all correlated effects are absorbed by the *other* mechanisms). Our upper estimate corresponds to the reduction when they are entered first (when all correlated effects are absorbed by the *given* mechanism). Recall that the gap that we aim to explain in reading comprehension is 0.97 and that the gap in mathematics is 1.05.

**Table 7**

**The relative strength of the transmission mechanisms: reduction in the size of the Roma coefficient in the test score regressions due to the variables corresponding to each mechanism**

	Reading		Mathematics	
	lower estimate	upper estimate	lower estimate	upper estimate
Health	0.01	0.10	0.03	0.11
Home environment	0.28	0.76	0.28	0.69
School	0.13	0.60	0.17	0.58

The lower estimate corresponds to the reduction in the Roma coefficient when the variables of the particular mechanism are entered last; the upper estimate corresponds to the reduction when they are entered first.

Although the range of the estimates is rather broad, the home learning environment and the likelihood of accessing to high-quality education appear to be very important. The results are consistent with the causal interpretation that the test scores of Roma students are worse *because* they have limited access to resources and activities that promote their skill development at home and *because* they have limited access to high-quality education services. Health appears to play a less important role in teenage test results; however, childhood health problems may affect later life outcomes<sup>6</sup> through other channels (Elo and Preston, 1992; Case, Lubotsky and Paxson, 2002; Case, Fertig and Paxson, 2005; Smith, 1999; 2009; Marmot and Wilkinson, 2006; Strauss and Thomas, 2008, chapter 4).

Having established the potential role of each mechanism, our next question is, to what degree do the Roma students' socio-economic disadvantages explain their deficits in home environment and access to educational services, and what is the potential role of ethnicity *per se*?

**Home environment and parenting**

As when we analyzed the determinants of the test score gap, not only in the neighborhood of the average values but also along the entire distribution of the family background scale, we examine the ethnic differences in the home environment, again throughout the distribution of family background. We use the previously introduced synthetic family background index for that purpose. Analogously to the previous analysis, we divide the range of the family background variable (the linear combination of family income, poverty, parental education and parental employment) into 10 equal intervals and estimate the mean values of the home environment variables for the Roma and non-Roma students. Similar to the previous analysis, we restrict the estimates for the Roma students to the 0–0.6 range, as the subsample is too small to be meaningful over values of 0.6.

We present our results in the following two figures. Social and ethnic differences in the frequency of bedtime storytelling to *kindergarten age* children, the number of books in the home and internet

<sup>6</sup> Such outcomes include adult health, mortality, employment and earnings.

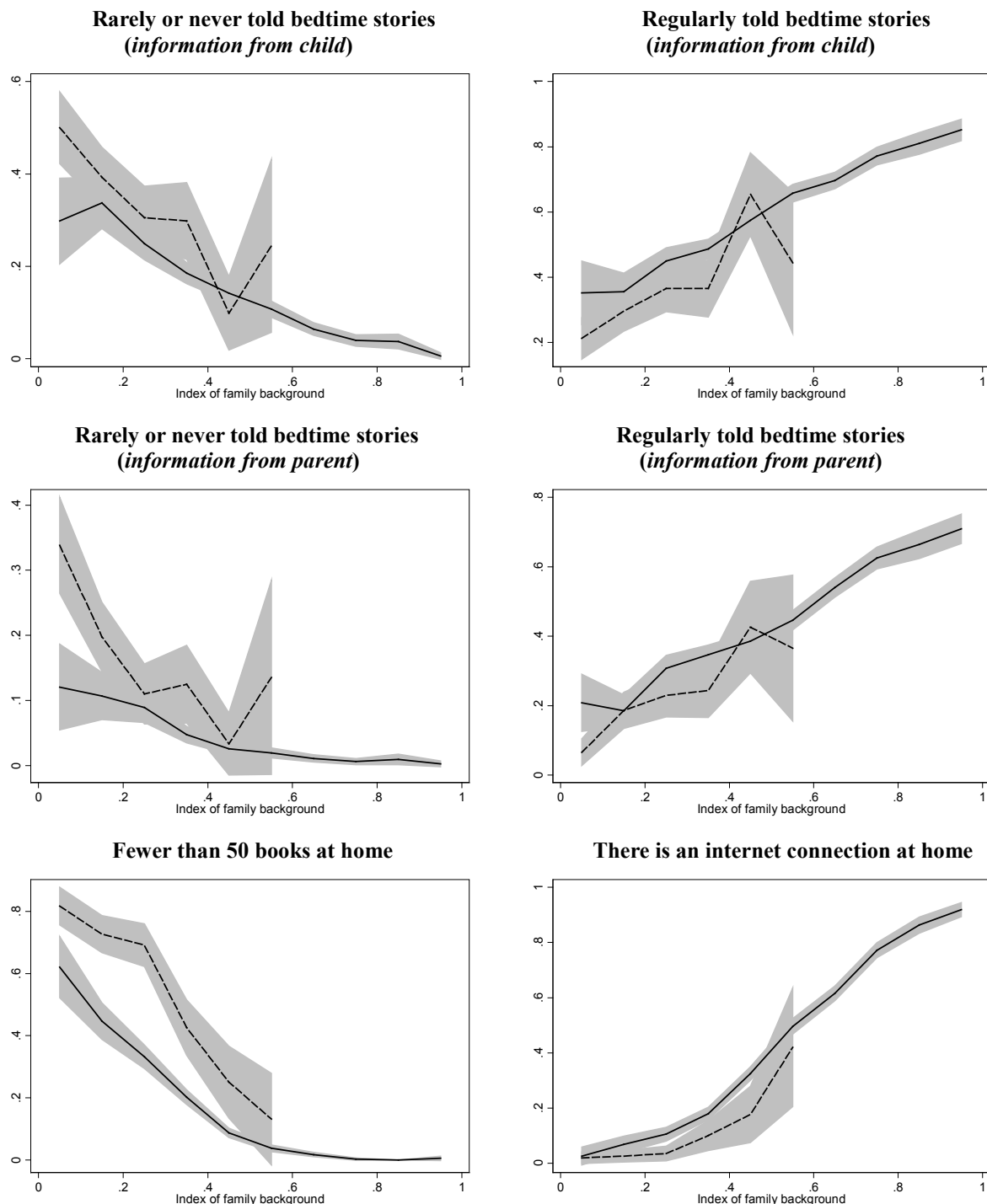


access in the home are reported in *Figure 6*, and the differences in the cognitive and emotional HOME index scores are presented in *Figure 7*. For expositional purposes, we omit the confidence intervals around the HOME index figures (they overlap across Roma and non-Roma, suggesting no significant differences).<sup>7</sup>

**Figure 6**

**The probability of bedtime storytelling, having no or very few books and having an internet connection at home as a function of the family background index**

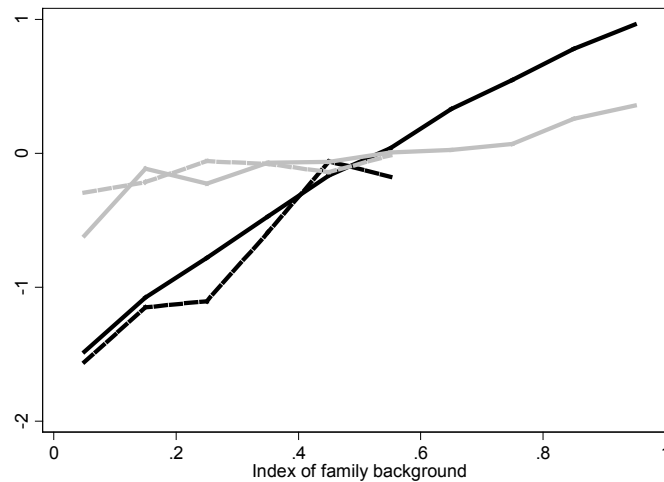
Solid lines: Non-Roma. Dashed lines: Roma. The gray zone indicates the 95 percent confidence intervals.



<sup>7</sup> As robustness checks, we estimated linear regressions with the family background variables entered separately. The results, shown in Table A8 in the Appendix, are very similar to the results in the figures below.

**Figure 7**  
**Family background and the cognitive and emotional HOME index**

Solid lines: Non-Roma; dashed lines: Roma.  
 Black lines: cognitive HOME index; gray lines: emotional HOME index



First, *Figures 6 and 7* reveal that most indicators are strongly related to the family background index within both ethnic groups. Only 20-30 percent of the most disadvantaged students were told bedtime stories regularly in early childhood, compared to 70-80 percent of children from the highest social status families. Of the poorest and least educated families, 70 percent have either no or very few books, which is true of none of the highest status families. Fewer than 5 percent of the poorest families had home internet access in 2006, compared to 90 percent of the wealthiest families. The difference in the cognitive HOME index (a comprehensive measure of the cognitive stimuli in the teenage home environment) between the two groups is a staggering 2.5 standard deviations. The exception is the emotional HOME index, which is weakly associated with family background.

Second, the figures reveal small and, in many cases, statistically insignificant ethnic differences in the home environment and parenting indicators between families with comparable family backgrounds. Statistically, no difference is found in storytelling or the cognitive and emotional HOME index graphs between the Roma and non-Roma students. Ethnic differences become small but remain statistically significant in terms of the number of books owned and internet access conditional on the family background index. However, these ethnic differences are smaller at higher levels of the family background index. If one were willing to extrapolate to the upper part of the distribution of family background, one could conclude that Roma students would have similar home environments to non-Roma students if they lived in similarly good circumstances.

The third interesting result is the flat profile of the emotional HOME index with respect to family background. The results indicate that high- and low-income families generally do not substantially differ in their capacity to provide emotional support to their children. This result is surprising, as the bottom third of society faces serious economic difficulties, and unemployment and economic hardship represent a major source of stress for families living in poor socio-economic circumstances. Parents living in poverty are nevertheless able to provide their children with nearly as much emotional support as parents of higher social status. Coupled with the insignificant ethnic differences conditional on family background, this result suggests that typical Roma families provide their children with the same level of emotional support as typical non-Roma families, even

though they face much more difficult economic conditions.<sup>8</sup>

We can only speculate about the reasons why children living in adverse circumstances have suboptimal access to the objects, activities and experiences that promote their skill development in their home environment. The most obvious cause is *income poverty*: low-income families are less able to afford the objects, tools and services that promote skill development than wealthier families. The role of income poverty is supported by recent studies from the United States (Duncan and Murnane, 2011b, p. 11; Kaushal, Magnuson and Waldfogel, 2011). Families that differ in parental education—and thus income—also differ in terms of *parental time use*. Less-educated parents are found to spend significantly less time with their children than more educated parents—even though they are less likely to be employed and spend more time at home on average (Sayer, Gauthier and Furstenberg, 2004, p. 1164; Guryan, Hurst and Kearney, 2008, p. 35; Ramey and Ramey, 2010, p. 137). Parental education is also associated with the quantity and quality of *parent-child interactions*. Less-educated parents speak with their children significantly less often, have less developed vocabulary and incorporate less encouragement and more discouragement in their parenting than more educated parents (Réger, 1990; Hoff-Ginsberg, 1991; Hart and Risley, 1995; Huttenlocher et al., 2002; Hoff, 2003; 2006; Phillips, 2011). Finally, disadvantaged children have more limited access to the objects, activities and experiences that promote their development than other children not only in their immediate family environment but also in their residential area (Neuman and Celano, 2001; Neuman, 1999; Neuman et al., 2001; Neuman and Celano, 2004).

### **Access to adequate education**

The second important mechanism behind the test score gap between Roma and non-Roma students is the Roma students' relatively limited access to adequate education. Ethnic inequalities in access are due to in part residential inequalities and in part selection mechanisms irrespective of residence. The majority of Roma students are educated in classrooms in which the sheer quantity of unresolved pedagogical problems makes it very difficult for teachers to teach well. To measure this, we combine the HLCS sample with the full 2006 NABC database. For every student in the HLCS sample, we calculated the percentage of the classmates whose reading test results were inadequate (levels 0 or 1; the maximum is 4; overall, 30 percent perform at this inadequate level). We then characterize the class of each student as *problematic* if the reading test results were inadequate for more than half of the student's classmates. As we argued previously, studying in problematic classes is detrimental to student development because the pedagogical difficulties lead to lower quality teaching. Moreover, these difficulties can adversely affect the quality of teachers through their self-selection, and direct peer effects may further hinder individual development.

We find substantial ethnic differences in the likelihood of studying in problematic classes. Of Roma

---

<sup>8</sup> Although emotional support is not strongly related to poverty, it is closely connected to family structure. In our sample, two-parent families are able to provide the highest levels of emotional support, and single mothers the lowest levels of emotional support. The difference between these two family types accounts for 70 percent of the standard deviation of the HOME emotional subscale scores. The comparable difference is much smaller in the case of the HOME cognitive subscale, barely exceeding 20 percent. It is important to note that the distribution of single parenthood and patchwork families is very similar across Roma and non-Roma households. Detailed results are available from the authors upon request.

8<sup>th</sup> graders, 58 percent are in problematic classes (in which over half of their classmates can be considered functionally illiterate), compared to 18 percent of non-Roma students. The raw ethnic difference is 40 percentage points. We estimated linear probability models to capture the ethnic difference while controlling for family background and home environment. Table 8 presents the results.

When controlling for the family background variables (which include residential information variables), we find that the ethnic difference declines substantially but remains non-negligible and statistically significant at 14 percentage points (see Table A9 in the Appendix). Thus, Roma children are 14 percentage points more likely to attend problematic classes than non-Roma children of similar family background. When we control for home environment and parenting variables in addition to the family background variables, the ethnic difference remains statistically significant at 12 percentage points. In conjunction, these results suggest that residential inequalities and self-selection by family background are responsible for the majority of the selection; however, schools are likely responsible for at least a quarter of the selection.

**Table 8**  
**The probability of attending a class that is problematic**  
(fraction of classmates with inadequate reading skills above 50 percent).  
Linear probability models. Number of observations: 9056

Roma	0.40 (0.022)**	0.14 (0.026)**	0.21 (0.025)**	0.12 (0.026)**
Family background	–	yes	–	yes
Home environment	–	–	yes	yes
Number of observations	9056	9056	9056	9056
R <sup>2</sup>	0.07	0.18	0.16	0.2

Standard errors in parentheses are clustered at the school level.

\*\* Significant at 1 %.

*Note:* see detailed results in Table A9 of the Appendix.

## Summary and policy recommendations

Using unique data from Hungary, this study measured the gap in standardized test scores between Roma and non-Roma 8<sup>th</sup> grade students and demonstrated that this gap is comparable to the size of the Black-White test score gap in the United States in the 1980s. The skills gap emerges at a very early age, before enrollment in elementary school, and that the differences measured at the end of elementary school continue to increase in secondary school.

Social differences (in income, parental education and place of residence) account for a large part of the test score gap. If the non-Roma students lived in socio-economic circumstances similar to those of Roma students, only a fraction of the gap would persist: one-fourth of the mathematics gap and one-fifth of the reading gap. Based on theoretical considerations and empirical results from the international literature, we examined the role of three mediating mechanisms by which these social differences could give rise to the ethnic test score gap: health, home environment and parenting and schools. We found that, together, these mechanisms completely explain the role of social differences in the test score gap and that they in themselves explain the entire gap in reading and 90 percent of

the gap in mathematics. Two of these mechanisms were particularly important: (1) home environment and parenting and (2) schools. We then demonstrated that ethnic differences in the home environment and parenting can be almost entirely explained by social differences (with the exception of books), especially in the middle income range (which is the highest end of the income range for Roma families). These factors are very strongly related to social differences, with the surprising exception of emotional support in families. We also found that the Roma students have substantial disadvantages with respect to access to adequate education. Roma students are 40 percentage points more likely to study in classrooms in which the majority of their peers have inadequate reading skills than non-Roma students. This increased likelihood is in large part due to residential and social disadvantages, but ethnicity remains a significant factor in the school system's selection mechanisms.

We conclude that the test score gap between Roma and non-Roma 8<sup>th</sup> grade students is primarily due to poverty and associated disadvantages at home and at school. Thus, aside from the phenomenon of school segregation, the causes of the achievement gap call for universal and color-blind policies instead of interventions targeting the Roma minority in particular. Policies that improve the long-run life chances of families with children in extreme poverty can result in substantial improvements in the children's skill development. Policies targeting the causal mechanisms directly are additional candidates.

Perhaps the most promising methods to prevent school failures are to provide children with an environment (objects, tools, activities, services) that facilitates their cognitive and language development and to promote complementary parenting methods (Herczog, 2008; Almond and Currie, 2011; Heckman, 2011). Unequal access to high-quality learning environments due to residential disadvantages and the selection mechanisms of the school system calls for additional policies aiming to improve and modernize the entire school system and incorporate pedagogical innovations to better integrate children from disadvantaged families, reduce school segregation and provide appropriate training and incentives for teachers that work in problematic educational environments

The skill development and school careers of disadvantaged children—including Roma children living in poverty—will largely depend on whether we prove capable of understanding and accepting evidence regarding the mediating mechanisms between poverty and low school achievement. This is what we must build on to shape social policy in a way that uses available resources as efficiently as possible to help these children and their families.

## References

- Akerlof, G. A. and Kranton, R. E. (2002), 'Identity and schooling. Some lessons for the economics of education', *Journal of Economic Literature*, 40(4), pp. 1167–1201.
- Almond, D. and Currie, J. (2011), 'Human capital development before age five', In Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, Vol. 4B. Elsevier, Amsterdam, pp. 1315–1486.
- Ammermueller, A. and Pischke, J-S. (2009), 'Peer effects in European primary schools', *Journal of Labor Economics*, 27(3), pp. 315–348.
- Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A. and Walters, C. R. (2010), 'Inputs and impacts in charter schools: KIPP Lynn', *American Economic Review*, 100(2), pp. 1–5.
- Barker, D. J. P. (1998), *Mothers, Babies and Health in Later Life*. Churchill Livingstone, Edinburgh.
- Becker, G. S. (1981a), 'The demand for children', In Becker, G. S., *A Treatise on the Family*, Harvard University Press, Cambridge, Mass. Enlarged edition: 1991, pp. 135–154.
- Becker, G. S. (1981b), 'Family background and the opportunities of children', In Becker, G. S., *A Treatise on the Family*, Harvard University Press, Cambridge, Mass. Enlarged edition: 1991, pp. 179–200.
- Becker, G. S. and Tomes, N. (1986), 'Human capital and the rise and fall of families', *Journal of Labour Economics*, 4(3), pp. S1–S39.
- Behrman, J. R. and Rosenzweig, M. R. (2004), 'Returns to birthweight', *Review of Economics and Statistics*, 86(2), pp. 586–601.
- Bishop, J. H., Bishop, M., Gelbwasser, L., Green, S. and Zuckerman, A. (2003), 'Nerds and freaks: a theory of student culture and norms', *Brookings Papers on Education Policy*, pp. 141–213.
- Black, S. E., Devereux, P. J. and Salvanes, K. G. (2007), 'From the cradle to the labor market. The effect of birth weight on adult outcomes', *Quarterly Journal of Economics*, 122(1), pp. 409–439.
- Bradley, R. H. et al. (2000): 'Measuring the HOME environments of children in early adolescence', *Journal of Research on Adolescence*, 10(3), pp. 247–288.
- Breslau, N. et al. (1994), 'A gradient relationship between low birth weight and IQ at age 6 years', *Archives of Pediatrics & Adolescent Medicine*, 148(4), pp. 377–383.
- Brooks-Gunn, J. and Markman, L. G. (2005), 'The contribution of parenting to ethnic and racial gaps in school readiness', *The Future of Children*, 15(1), pp. 139–168.
- Case, A., Fertig, A. and Paxson, C. (2005), 'The lasting impact of childhood health and circumstance', *Journal of Health Economics*, 24(2), pp. 365–389.
- Case, A., Lubotsky, D. and Paxson, C. (2002), 'Economic status and health in childhood: the origins of the gradient', *American Economic Review*, 92(5), 1308–1334.
- Chetty, R., Friedman, J. N. and Rockoff, J. E. (2011), 'The long-term impacts of teachers: teacher value-added and student outcomes in adulthood', *NBER Working Paper No. 17699*.
- Clotfelter, C. T., Ladd, H. F. and Vigdor, J. (2005), "Who teaches whom? Race and the distribution of novice teachers", *Economics of Education Review*, 24(4), pp. 377–392.
- Conger, R. D., Conger, K. J., Elder, G. H., Lorenz, F. O., Simons, R. L. and Whitbeck, L. B. (1992), "A family process model of economic hardship and adjustment of early adolescent boys", *Child Development*, 63(3), pp. 523–541.

- Conger, R. D., Conger, K. J., Elder, G. H., Lorenz, F. O., Simons, R. L. and Whitbeck, L. B. (1993), 'A family economic stress and adjustment of early adolescent girls', *Development Psychology*, 29(2), pp. 206–219.
- Cramer, J. C. (1995), 'Racial and ethnic differences in birthweight: the role of income and financial assistance', *Demography*, 32(2), pp. 231–247.
- Crane, J. (1996), 'Effects of home environment, SES, and maternal test scores on mathematics achievement', *Journal of Educational Research*, 89(5), pp. 305–314.
- Currie, J. (2011), 'Inequality at birth: some causes and consequences', *American Economic Review*, 101(3), pp. 1–22.
- Currie, J. (2005), 'Health disparities and gaps in school readiness', *The Future of Children*, 15(1), pp. 117–138.
- Currie, J. (2009), 'Healthy, wealthy, and wise', *Journal of Economic Literature*, 47(1), pp. 87–122.
- Currie, J. and Hyson, R. (1999), 'Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight', *American Economic Review*, 89(2), pp. 245–250.
- Currie, J. and Stabile, M. (2003), 'Socioeconomic status and child health. Why is the relationship stronger for older children?', *American Economic Review*, 93(5), pp. 1813–1823.
- Csapó, B. (2007), 'Hosszmetszeti felmérések iskolai kontextusban' [Longitudinal surveys in educational context], *Magyar Pedagógia*, 107(4), pp. 321–355.
- Dickinson, D. K. and Tabors, P. O. (2001), *Beginning Literacy with Language*, Paul H. Brookes, Baltimore.
- Dobbie, W. and Fryer, R. G. (2011), 'Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone', *American Economic Journal: Applied Economics*, 3(3), pp. 158–187.
- Duncan, G. J. and Murnane, R. J. (eds.) (2011a), *Wither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, Russel Sage Foundation–Spencer Foundation, New York, Chicago.
- Duncan, G. J. and Murnane, R. J. (2011b), 'Introduction: the American dream, then and now', In Duncan, G. J. and Murnane, R. J. (eds) (2011a), pp. 3–23.
- Elder, G. H. Jr. (1974), *Children of the Great Depression. Social Change in Life Experience*, The University of Chicago Press, Chicago, London.
- Elo, I. T. and Preston, S. H. (1992), 'Effects of early-life conditions on adult mortality: a review', *Population Index*, 58(2), pp. 186–212.
- Ferguson, R. F. (1998), 'Can schools narrow the black-white test score gap?', In Jencks, C. and Phillips, M. (eds), *The Black-White Test Score Gap*, Brookings Institution Press, Washington, D.C., pp. 318–374.
- FRA-UNDP (2012), *The situation of Roma in 11 EU Member States*, European Agency of Fundamental Rights and UNDP.
- Fryer, R. G. and Torelli, P. (2010), 'An empirical analysis of „acting white”', *Journal of Public Economics*, 94(5–6), pp. 380–396.
- Gould, E. D. and Simhon, A. (2011), 'Does quality time produce quality children?', *IZA Discussion Paper* No. 5487.
- Guo, G. and Harris, K. M. (2000), 'The mechanisms mediating the effects of poverty on children's intellectual development', *Demography*, 37(4), pp. 431–447.

- Guryan, J., Hearst, E. and Kearney, M. (2008), 'Parental education and parental time with children', *Journal of Economic Perspectives*, 22(3), pp. 23–46.
- Hack, M., Klein, N. K. and Taylor, H. G. (1995), 'Long-term developmental outcomes of low-birth weight infants', *The Future of Children*, 5(1), pp. 176–196.
- Hanushek, E. A., Kain, J. F., Markman, J. M. and Rivkin, S. G. (2003), 'Does peer ability affects student achievement?', *Journal of Applied Econometrics*, 18(5), pp. 527– 544.
- Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2004), 'Why public schools lose teachers?', *Journal of Human Resources*, 39(2), pp. 326–354.
- Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2009), 'Brown v. Board of Education: the complex effects of school racial composition on achievement', *Journal of Labor Economics*, 27(3), pp. 349–383.
- Hart, B. and Risley, T. (1995), *Meaningful Differences in the Everyday Experience of Young American Children*, Brookes, Baltimore.
- Haveman, R. and Wolfe, B. (1995), 'The determinants of children's attainments: a review of methods and findings', *Journal of Economic Literature*, 33(4), pp. 1829–1878.
- Heath, S. B. (1984), *Ways with Words: Language, Life, and Work in Communities and Classrooms*, Cambridge University Press.
- Heckman, J. J. (2011), 'The American family in Black & White: a post-racial strategy for improving skills to promote equality', *Daedalus*, 140(2), pp. 70–89.
- Herczog, M. (2008), 'Encouraging early child development' In Fazekas, K., Köllő, J. and Varga, J. (eds), *Green Book for the Renewal of Public Education in Hungary*, Ecostat, Budapest, pp. 33–52.
- Hoff, E. (2003), 'The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech', *Child Development*, 74(5), pp. 1368–1378.
- Hoff, E. (2006), 'How social contexts support and shape language development', *Developmental Review*, 26(1), pp. 55–88.
- Hoff-Ginsberg, E. (1991), 'Mother-child conversation in different social classes and communicative settings', *Child Development*, 62(4), pp. 782–796.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E. and Levine, S. (2002), 'Language input and child syntax', *Cognitive Psychology*, 45(3), pp. 337–374.
- Jackson, C. K. (2009), 'Student demographics, teacher sorting, and teacher quality: evidence from the end of school desegregation', *Journal of Labor Economics*, 27(2), pp. 213–256.
- Kalil, A. and DeLeire, T. (2004), *Family Investments in Children's Potential. Resources and Parenting Behaviors that Promote Success*, Lawrence Erlbaum, Mahwah (NJ).
- Kaushal, N., Magnuson, K. and Waldfogel, J. (2011), 'How is family income related to investments in children's learning?' In Duncan, G. J. and Murnane, R. J. (eds) (2011a), pp. 187–205.
- Kertesi, G. and Kézdi, G. (2010), 'Roma employment in Hungary after the post-communist transition', *Economics of Transition*, 19(3), pp. 563-610.
- Kertesi, G. and Kézdi, G. (2011), 'The Roma/non-Roma test score gap in Hungary', *American Economic Review*, 101(3), pp. 519–525.
- Kézdi, G. and Surányi, É. (2008), *Egy sikeres iskolai integrációs program tapasztalatai* [Experiences of a Successful School Integration Program], Educatio, Budapest.
- Köllő, J. (1997), 'A napi ingázás feltételei és a helyi munkanélküliség Magyarországon: számítások és számpéldák' [Conditions of daily commuting and local unemployment in Hungary], *Esély*, 8(2), pp. 33–61.



- Lee, V. and Burkam, D. (2002), *Inequality at the Starting Gate. Social Background Differences in Achievement as Children Begin School*, Economic Policy Institute, Washington, DC.
- Leibowitz, A. (1974), 'Home investment in children', *Journal of Political Economy*, 82(2), pp. S111–S131.
- Lempers, J. D., Clark-Lempers, D. and Simons, R. L. (1989), 'Economic hardship, parenting, and distress in adolescence', *Child Development*, 60(10), pp. 25–39.
- Linver, M. R., Brooks-Gunn, J. and Kohen, D. (2002), 'Family processes as pathways from income to young children's development', *Development Psychology*, 38(5), pp. 719–734.
- Marmot, M. and Wilkinson, R. G. (2006), *Social Determinants of Health*, Oxford University Press, Oxford.
- Mayer, S. E. (1997), *What Money Can't Buy. Family Income and Children's Life Chances*, Harvard University Press, Cambridge, Mass.
- McLoyd, V. (1990), 'The impact of economic hardship on black families and children: psychological distress, parenting, and socioemotional development', *Child Development*, 61(2), pp. 311–346.
- Meara, E. (2001), 'Why is health related to socioeconomic status? The case of pregnancy and low birth weight', *NBER Working Paper No. 8231*.
- Mott, F. L. (2004), 'The utility of the HOME-SF scale for child development research in a large longitudinal survey: the National Longitudinal Survey of Youth 1979 cohort', *Parenting: Science & Practice*, 4(2–3), pp. 259–270.
- Mulligan, C. B. (1997), *Parental Priorities and Economic Inequality*, The University of Chicago Press, Chicago–London.
- Neuman, S. B. (1996), 'Children engaging in storybook reading', *Early Childhood Research Quarterly*, 11(4), pp. 495–513.
- Neuman, S. B. (1999), 'Books make a difference. A study of access to literacy', *Reading Research Quarterly*, 34(3), pp. 286–311.
- Neuman, S. B. (2006), 'The knowledge gap', In Dickinson, D. K. and Neuman, S. B. (eds), *Handbook of Early Literacy Research*, Vol 2. The Guilford Press, New York–London. pp. 29–40.
- Neuman, S. B. and Celano, D. C. (2001), 'Access to print in low-income and middle-income communities: an ecological study of four neighborhoods', *Reading Research Quarterly*, 36(1), pp. 8–26.
- Neuman, S. B.–Celano, D. C. (2004), 'Save the libraries!', *Educational Leadership*, 61(6), pp. 82–85.
- Neuman, S. B., Celano, D. C., Greco, A. N. and Shue, P. (2001), *Access for All. Closing the Book Gap for Children in Early Education*, International Reading Association, Newark, DE.
- Oreopoulos, P., Stabile, M., Walld, R. and Roos, L. L. (2008), 'Short-, medium-, and long-term consequences of poor infant health', *Journal of Human Resources*, 43(1), pp. 88–138.
- Palloni, A., Milesi, C., White, R. G. and Turner, A. (2009), 'Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials', *Social Science & Medicine*, 68(9), pp. 1574–1582.
- Paul, A. M. (2010), *Origins. How the Nine Months Before Birth Shape the Rest of Our Lives*, Free Press, New York.
- Phillips, M. (2011), 'Parenting, time use, and disparities in academic outcomes', In Duncan, G. J. and Murnane, R. J. (eds) (2011a), pp. 207–228.

- Raikes, H. et al. (2006), 'Mother-child bookreading in low-income families: correlates and outcomes during the first three years of life', *Child Development*, 77(4), pp. 924–953.
- Ramey, G. and Ramey, V. A. (2010), 'The rug rat race', *Brookings Papers on Economic Activity*, 41(1), pp. 129–199.
- Réger, Z (1990), *Utak a nyelvhez: nyelvi szocializáció – nyelvi hátrány* [Pathways to the Language: Socialization and Language Disadvantage], Akadémiai Kiadó, Budapest.
- Reichman, N. E. (2005), 'Low birth weigh and school readiness', *The Future of Children*, 15(1), pp. 91–116.
- Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *Econometrica*, 73(2), pp. 417–458.
- Rosenzweig, M. R. and Schultz, T. P. (1982): 'The behavior of mothers as inputs to child health: the determinants of birth weight, gestation, and rate of fetal growth', In Fuchs, V. R. (ed), *Economic Aspects of Health*, University of Chicago Press, NBER, pp. 53–92.
- Sayer, L. C., Gauthier, A. H. and Furstenberg, F. F. (2004), 'Educational differences in parents' time with children: cross-national variations', *Journal of Marriage and Family*, 66(5), pp. 1152–1169.
- Schacter, J. and Thum, Y. M. (2004), 'Paying for high and low quality teaching', *Economics of Education Review*, 23(4), pp. 411–430.
- Schonkoff, J. P. and Phillips, D. A. (2004), *From Neurons to Neighborhoods. The Science of Early Childhood Development*, National Academy Press, Washington, D.C.
- Sénéchal, M., LeFevre, J-A., Hudson, E. and Lawson, E. P. (1996), 'Knowledge of storybooks as a predictor of young children's vocabulary' *Journal of Educational Psychology*, 88(3), pp. 520–536.
- Smith, J. P. (1999), 'Healthy bodies and thick wallets: the dual relation between health and economic status', *Journal of Economic Perspectives*, 13(2), pp. 145–166.
- Smith, J. P. (2009), 'Reconstructing childhood health histories', *Demography*, 46(2), pp. 387–403.
- Strauss, J. and Thomas, D. (2008), 'Health over the life course' In Schultz, T. P. and Strauss, J. (eds), *Handbook of Development Economics*, Vol. 4. Elsevier, Amsterdam, pp. 3376–3474.
- Todd, P. E. and Wolpin, K. I. (2007), 'The production of cognitive achievement in children: home, school, and racial test score gaps', *Journal of Human Capital*, 1(1), pp. 91–136.
- Wenglinsky, H. (2001), 'Teacher classroom practices and student performance. How schools can make a difference', Research Report RR-01-19, Educational Testing Service, Statistics & Research Division, Princeton.

## Appendix

Table A1. Measurement of Roma ethnicity in the Hungarian Life Course Survey

Ethnic identity	Mother	Father	Mother or father
Chose the Roma identity as his/her first choice in wave 1	2.4	2.6	3.0
Chose the Roma identity as his/her second choice in wave 1	3.4	3.6	3.4
Only chose the Roma identity in wave 2, there as his/her first choice	0.5	0.5	0.5
Only chose the Roma identity in wave 2, there as his/her second choice	0.9	0.8	1.0
Did not choose the Roma identity in either case	91.1	74.4	92.2
No parent, or all parental nationality-ethnicity data are missing	1.7	18.2	0.0
Total	100.0	100.0	100.0

Table A2. Sample selection of the Hungarian Life Course Survey for our analysis

	Number of observations	Standardized test score average <sup>a</sup>		Proportion of students whose mother	
		Reading	Mathematics	has completed no more than 8 years of school	has completed higher education
Based on National Assessment of Basic Competencies data on 8 <sup>th</sup> graders in 2006					
Total students	113,092	-	-	-	-
Students who completed the reading test	109,906	-0.08	-	-	-
Students who completed the mathematics test	104,566	-	-0.06	-	-
Students who completed the reading and mathematics tests	104,533	-0.03	-0.06	-	-
Students with test scores and family background data	88,175	-0.01	-0.04	0.18	0.21
Among them: students whose families have agreed to participate in the Hungarian Life Course Survey	37,027	-0.14	-0.09	0.24	0.19
Based on Hungarian Life Course Survey data					
Sample in the first wave <sup>b</sup>	10,022	-0.11	-0.05	0.21	0.20
Sample in the second wave <sup>b</sup>	9,300	-0.10	-0.04	0.21	0.20
The sample that forms the basis of our estimates <sup>b</sup>	9,056	-0.09	-0.03	0.20	0.20

<sup>a</sup> Values standardized for the average and standard deviation of national test scores (theoretical average 0, theoretical standard distribution 1; real averages may differ slightly as not all students' results were used)

<sup>b</sup> The statistics drawn from the HLCS are weighted values (using the sampling weights)

Table A3. Summary statistics of the family background variables, Roma and non-Roma subsamples (weighted averages and standard deviations)

	Roma subsample		Non-Roma subsample	
	average	standard deviation	average	standard deviation
Biological mother in household	0.96	0.20	0.97	0.18
Non-biological mother in household	0.03	0.17	0.01	0.11
Biological father in household	0.78	0.41	0.72	0.45
Non-biological father in household	0.06	0.24	0.09	0.28
Mother's education: grades 0-8	0.79	0.41	0.15	0.36
Mother's education: vocational school	0.15	0.36	0.25	0.43
Mother's education: high school diploma	0.04	0.20	0.36	0.48
Mother's education: higher education	0.01	0.08	0.22	0.41
Father's education: grades 0-8	0.54	0.50	0.08	0.27
Father's education: vocational school	0.27	0.44	0.37	0.48
Father's education: high school diploma	0.03	0.18	0.21	0.41
Father's education: higher education	0.00	0.06	0.14	0.35
Mother employed	0.24	0.43	0.70	0.00
Father employed	0.35	0.48	0.66	0.47
Proportion of years mother employed while child was age 0-14	0.30	0.35	0.64	0.32
Proportion of years father employed while child was age 0-14	0.52	0.45	0.73	0.43
Logarithm of family income	11.68	0.46	12.03	0.46
Logarithm of household size	1.58	0.35	1.39	0.29
Number of unemployed adults	1.39	0.99	0.67	0.81
Size of apartment, m2 per person	17.55	9.62	23.57	10.16
Number of rooms per person	0.55	0.25	0.79	0.29
Bathroom in apartment	0.75	0.43	0.97	0.17
No money for food	0.23	0.42	0.05	0.21
No money for heating	0.35	0.48	0.12	0.32
Received regularized child-rearing assistance	0.67	0.47	0.22	0.42
Free lunch in 8th grade	0.17	0.38	0.08	0.27
Free textbooks in 8th grade	0.87	0.33	0.56	0.50
Mother's education - data missing	0.02	0.13	0.02	0.14
Father's education - data missing	0.15	0.36	0.20	0.40
Family income - data missing	0.06	0.25	0.10	0.30
Size of apartment - data missing	0.05	0.21	0.01	0.11
Number of rooms - data missing	0.00	0.05	0.00	0.06
Bathroom - data missing	0.00	0.05	0.00	0.04
Poverty indicator - data missing	0.01	0.08	0.01	0.08
Region: Central	0.07	0.26	0.22	0.41
Region: Central Transdanubia	0.06	0.23	0.12	0.33
Region: Western Transdanubia	0.06	0.24	0.13	0.34
Region: Southern Transdanubia	0.25	0.43	0.12	0.32
Region: Northern Hungary	0.31	0.46	0.11	0.32
Region: Northern Great Plain	0.18	0.38	0.16	0.37
Region: Southern Great Plain	0.08	0.27	0.14	0.34
Budapest	0.05	0.21	0.13	0.34
County seat	0.07	0.26	0.17	0.38
Other city	0.32	0.47	0.35	0.48
Settlement	0.56	0.50	0.34	0.47
Remote settlement	0.18	0.39	0.12	0.32
Number of observations		848		8208

Table A4. Detailed OLS regression estimates for *Table 4* (dependent variables: test scores, independent variables: family background)

	Dependent variable			
	Reading test scores		Mathematics test scores	
Roma	-0.97 (0.053)**	-0.23 (0.055)**	-1.047 (0.048)**	-0.324 (0.050)**
Biological mother in household		0.05 (0.231)		-0.048 (0.253)
Non-biological mother in household		-0.19 (0.240)		-0.218 (0.266)
Biological father in household		0.01 (0.389)		-0.176 (0.217)
Non-biological father in household		-0.03 (0.389)		-0.261 (0.219)
Mother's education: grades 0-8		-0.67 (0.048)**		-0.659 (0.050)**
Mother's education: vocational school		-0.57 (0.038)**		-0.527 (0.042)**
Mother's education: higher education		-0.26 (0.033)**		-0.223 (0.038)**
Father's education: grades 0-8		-0.62 (0.053)**		-0.708 (0.061)**
Father's education: vocational school		-0.43 (0.040)**		-0.54 (0.047)**
Father's education: high school diploma		-0.25 (0.039)**		-0.265 (0.047)**
Mother employed		-0.02 (0.035)		-0.008 (0.037)
Father employed		0.03 (0.041)		-0.007 (0.042)
Proportion of years mother employed while child was age 0-14		-0.01 (0.044)		-0.007 (0.050)
Proportion of years father employed while child was age 0-14		0.19 (0.051)**		0.117 (0.057)*
Logarithm of family income		0.00 (0.028)		0.047 (0.031)
Logarithm of household size		-0.05 (0.055)		-0.02 (0.062)
Number of unemployed adults		-0.03 (0.018)		-0.02 (0.019)
Size of apartment, m <sup>2</sup> per person		0.00 (0.002)		0.001 (0.002)
Number of rooms per person		0.23 (0.057)**		0.227 (0.065)**
Bathroom in apartment		0.14 (0.062)*		0.133 (0.062)*
No money for food		-0.20 (0.050)**		-0.153 (0.052)**
No money for heating		-0.08 (0.036)*		-0.058 (0.037)
Received regularized child-rearing assistance		0.04 (0.031)		0 (0.032)
Free lunch in 8th grade		-0.16		-0.098

		(0.043)**		(0.049)*
Free textbooks in 8th grade		-0.09		-0.026
		(0.026)**		(0.029)
Mother's education - data missing		-0.67		-0.698
		(0.220)**		(0.240)**
Father's education - data missing		-0.21		-0.594
		(0.389)		(0.220)**
Family income - data missing		-0.02		-0.036
		(0.034)		(0.036)
Size of apartment - data missing		-0.14		-0.155
		(0.104)		(0.104)
Number of rooms - data missing		0.03		0.277
		(0.162)		(0.241)
Bathroom - data missing		-0.13		0.19
		(0.171)		(0.184)
Poverty indicator - data missing		0.10		0.102
		(0.116)		(0.130)
Region: Central		-0.01		-0.077
		(0.056)		(0.058)
Region: Central Transdanubia		-0.04		-0.02
		(0.050)		(0.062)
Region: Western Transdanubia		-0.01		0.032
		(0.048)		(0.058)
Region: Southern Transdanubia		0.02		0.038
		(0.051)		(0.060)
Region: Northern Hungary		-0.08		-0.062
		(0.050)		(0.056)
Region: Northern Great Plain		-0.07		-0.072
		(0.046)		(0.054)
Budapest		0.19		0.212
		(0.060)**		(0.061)**
County seat		0.15		0.165
		(0.038)**		(0.044)**
Other city		0.04		0.044
		(0.030)		(0.034)
Remote settlement		0.04		0.04
		(0.040)		(0.043)
Constant	-0.02	0.22	0.044	0.054
	(0.017)	(0.544)	(0.019)*	(0.394)
Number of observations	9056	9056	8335	8335
R <sup>2</sup>	0.06	0.27	0.07	0.27

Robust standard errors clustered by school in parentheses

\*Significant at 5% level; \*\*Significant at 1% level

Table A5. Summary statistics of the health and home environment variables, Roma and non-Roma subsamples (weighted averages and standard deviations)

	Roma subsample		Non-Roma subsample	
	average	standard deviation	average	standard deviation
Low birth weight	0.17	0.38	0.07	0.25
Poor health (self-evaluation)	0.17	0.37	0.09	0.28
Weight - data missing	0.01	0.10	0.00	0.06
Health - data missing	0.01	0.10	0.01	0.09
Seldom or never told bedtime stories (child's response)	0.34	0.48	0.11	0.31
Often told bedtime stories (child's response)	0.35	0.48	0.65	0.48
Seldom or never told bedtime stories (parent's response)	0.18	0.38	0.03	0.16
Often told bedtime stories (parent's response)	0.21	0.41	0.48	0.50
Seldom went hiking with parents (child's response)	0.76	0.43	0.44	0.50
Cognitive HOME index	-1.03	0.98	0.09	0.94
Emotional HOME index	-0.17	0.98	0.02	0.98
Storytelling variable missing	0.05	0.22	0.04	0.20
Cognitive HOME variable missing	0.02	0.13	0.01	0.11
Emotional HOME variable missing	0.04	0.19	0.02	0.15
Number of books less than 50	0.64	0,48	0,09	0,28
Number of books around 50	0,16	0,37	0,11	0,32
Number of books: 50-150	0,11	0,31	0,23	0,42
Number of books: 150-300	0.04	0.20	0.20	0.40
Number of books: 300-600	0.02	0.15	0.17	0.37
Number of books: 600-1000	0.01	0.09	0.09	0.28
Number of books: more than 1000	0.01	0.10	0.11	0.31
Internet connection at home	0.07	0.25	0.51	0.50
Number of books - data missing	0.00	0.07	0.01	0.08
Internet connection - data missing	0.00	0.06	0.00	0.04
Number of observations	848		8208	

Table A6. Items in the short form of the young adolescent HOME scale (HOME-SF)<sup>a</sup>

	Survey question	Interviewer's observation
<b>COGNITIVE SUBSCALE</b>		
Has the child more than 20 books? (y/n)	X	
Is there a musical instrument? (y/n)	X	
Does the family get a daily newspaper? (y/n)	X	
Does the child read every week for enjoyment? (y/n)	X	
Does the family encourage the child to start and keep hobbies? (y/n)	X	
Does the child get special lessons? (y/n)	X	
Has the child been in museum last year with any family member? (y/n)	X	
Has the child been in musical or theatrical performance last year with any family member? (y/n)	X	
When the family watches TV together, do the parents discuss TV program with the child? (y/n)	X	
Is the interior of the home dark and perceptually monotonous? (y/n)		X
Are all visible rooms of the apartment reasonably clean? (y/n)		X
Are all visible rooms of the apartment minimally cluttered? (y/n)		X
Has the building potentially dangerous or health hazards? (y/n)		X
<b>EMOTIONAL SUBSCALE</b>		
How often is the child expected to clean his/her room?	X	
How often is the child expected to pick up after himself/herself?	X	
How often is the child expected to help manage his/her own time (get up on time, be ready for school)?	X	
How often does the whole family get together with relatives or friends?	X	
How often does the child spend time with his/her father?	X	
How often does the child spend time with his/her father in outdoor activities?	X	
How often does the child eat a meal with both mother and father?	X	
Sometimes children get so angry at their parents that they say things like „I hate you” or swear in temper tantrum. In this case would you spank your child? (y/n)	X	
Did you spank your child last week because of bad behaviour? (y/n)	X	
Mother/Guardian encouraged the child to contribute to the conversation with the interviewer. (y/n)		X
Mother/Guardian answered the child's questions or requests verbally during the interview. (y/n)		X
Mother/Guardian conversed with the child during the interview (excluding scolding or suspicious comments). (y/n)		X
Mother/Guardian introduces the interviewer to the child by name. (y/n)		X
Mother/Guardian's voice conveyed positive feeling about the child during the interview. (y/n)		X

<sup>a</sup> <http://www.bls.gov/nls/y79cyaguide/2002/y79chya20gac.pdf>, Appendix A.



Table A7. Detailed OLS regression estimates for *Table 6* (dependent variables: test scores, independent variables: health, home environment, school/class fixed effects, family background)

	Dependent variable					
	Reading test scores			Mathematics test scores		
Roma	-0.97 (0.053)**	-0.07 (0.072)	-0.05 (0.072)	-1.05 (0.048)**	-0.18 (0.066)**	-0.15 (0.067)*
Low birth weight		-0.09 (0.053)	-0.08 (0.052)		-0.18 (0.052)**	-0.16 (0.052)**
Poor health (self-evaluation)		-0.14 (0.049)**	-0.12 (0.049)*		-0.19 (0.056)**	-0.17 (0.056)**
Weight - data missing		-0.37 (0.213)	-0.34 (0.208)		-0.24 (0.196)	-0.18 (0.179)
Health - data missing		0.04 (0.136)	0.07 (0.134)		-0.02 (0.152)	0.00 (0.157)
Seldom or never told bedtime stories (child's response)		0.00 (0.054)	0.01 (0.054)		0.02 (0.053)	0.03 (0.054)
Often told bedtime stories (child's response)		0.10 (0.039)*	0.09 (0.038)*		0.06 (0.039)	0.05 (0.039)
Seldom or never told bedtime stories (parent's response)		-0.05 (0.077)	-0.07 (0.076)		-0.05 (0.072)	-0.05 (0.072)
Often told bedtime stories (parent's response)		0.08 (0.033)*	0.06 (0.033)		0.06 (0.036)	0.05 (0.035)
Seldom went hiking with parents (child's response)		0.01 (0.035)	0.02 (0.035)		-0.04 (0.036)	-0.02 (0.036)
Cognitive HOME index		0.18 (0.021)**	0.16 (0.022)**		0.14 (0.022)**	0.10 (0.023)**
Emotional HOME index		-0.03 (0.018)	-0.03 (0.019)		-0.04 (0.020)*	-0.04 (0.022)*
Storytelling variable missing		0.05 (0.082)	0.04 (0.082)		0.04 (0.090)	0.04 (0.088)
Cognitive HOME variable missing		0.00 (0.147)	-0.02 (0.151)		-0.17 (0.132)	-0.17 (0.131)
Emotional HOME variable missing		0.14 (0.118)	0.12 (0.120)		0.02 (0.100)	-0.01 (0.100)
Number of books less than 50		-0.48 (0.073)**	-0.42 (0.076)**		-0.39 (0.087)**	-0.27 (0.087)**
Number of books around 50		-0.36 (0.074)**	-0.29 (0.075)**		-0.34 (0.081)**	-0.21 (0.081)**
Number of books: 50-150		-0.29 (0.061)**	-0.24 (0.063)**		-0.23 (0.072)**	-0.14 (0.072)*
Number of books: 150-300		-0.16 (0.060)**	-0.11 (0.062)		-0.08 (0.073)	-0.01 (0.073)
Number of books: 300-600		-0.13 (0.061)*	-0.10 (0.062)		-0.09 (0.069)	-0.05 (0.068)
Number of books: 600-1000		-0.14 (0.071)*	-0.12 (0.071)		-0.10 (0.080)	-0.09 (0.080)
Internet connection at home		0.18 (0.037)**	0.15 (0.039)**		0.27 (0.039)**	0.23 (0.040)**
Number of books - data missing		-0.24 (0.170)	-0.18 (0.183)		-0.15 (0.242)	-0.10 (0.246)
Internet connection - data missing		-0.11 (0.215)	-0.16 (0.208)		-0.07 (0.222)	-0.27 (0.208)

Biological mother in household	-0.31 (0.335)	-0.05 (0.321)
Non-biological mother in household	-0.37 (0.337)	-0.03 (0.328)
Biological father in household	0.12 (0.482)	-0.58 (0.563)
Non-biological father in household	0.18 (0.482)	-0.59 (0.562)
Mother's education: grades 0-8	-0.12 (0.068)	-0.22 (0.071)**
Mother's education: vocational school	-0.18 (0.060)**	-0.22 (0.062)**
Mother's education: higher education	-0.06 (0.052)	-0.10 (0.055)
Father's education: grades 0-8	-0.21 (0.076)**	-0.27 (0.086)**
Father's education: vocational school	-0.16 (0.059)**	-0.20 (0.068)**
Father's education: high school diploma	-0.10 (0.059)	-0.09 (0.070)
Mother employed	0.01 (0.046)	0.03 (0.048)
Father employed	0.03 (0.052)	-0.04 (0.056)
Proportion of years mother employed while child was age 0-14	-0.11 (0.061)	-0.08 (0.063)
Proportion of years father employed while child was age 0-14	0.10 (0.071)	0.16 (0.074)*
Logarithm of family income	-0.03 (0.040)	0.01 (0.043)
Logarithm of household size	-0.10 (0.082)	-0.11 (0.082)
Number of unemployed adults	-0.03 (0.027)	-0.03 (0.027)
Size of apartment, m <sup>2</sup> per person	0.00 (0.002)	0.00 (0.002)
Number of rooms per person	-0.11 (0.080)	-0.07 (0.091)
Bathroom in apartment	-0.05 (0.077)	-0.02 (0.071)
No money for food	-0.03 (0.064)	-0.04 (0.061)
No money for heating	0.00 (0.048)	0.02 (0.050)
Received regularized child-rearing assistance	0.07 (0.044)	0.04 (0.047)
Free lunch in 8th grade	-0.12 (0.064)	-0.13 (0.062)*
Free textbooks in 8th grade	-0.06 (0.036)	0.03 (0.039)
Mother's education - data missing	-0.50 (0.319)	-0.40 (0.310)
Father's education - data missing	0.08 (0.484)	-0.69 (0.564)
Family income - data missing	-0.05 (0.049)	-0.08 (0.057)
Size of apartment - data missing	-0.05 (0.133)	-0.07 (0.119)

Number of rooms - data missing			0.20 (0.190)			0.53 (0.221)*
Bathroom - data missing			-0.25 (0.272)			0.19 (0.228)
Poverty indicator - data missing			-0.13 (0.159)			0.01 (0.195)
Region: Central Transdanubia			-0.49 (0.351)			0.24 (0.171)
Region: Central Transdanubia			0.63 (0.586)			0.67 (0.430)
Region: Western Transdanubia			-0.64 (0.551)			0.77 (0.359)*
Region: Southern Transdanubia			-1.35 (0.725)			-0.34 (0.484)
Region: Northern Hungary			-0.33 (0.514)			-0.05 (0.741)
Region: Northern Great Plain			-0.32 (0.445)			0.05 (0.703)
Budapest			-0.01 (0.184)			-0.06 (0.200)
County seat			0.05 (0.094)			-0.04 (0.119)
Other city			-0.08 (0.089)			-0.06 (0.098)
Remote settlement			0.09 (0.080)			0.09 (0.074)
Constant	-0.02 (0.017)		1.34 (0.854)	0.04 (0.019)*		0.78 (0.885)
Number of observations	9056	9056	9056	8335	8335	8335
R <sup>2</sup>	0.06	0.67	0.68	0.07	0.68	0.69

Robust standard errors clustered by school in parentheses

\*Significant at 5% level; \*\*Significant at 1% level

Table A8. The raw and corrected ethnic gap in the indicators of the home environment

Dependent variable	Roma coefficient	Standard error	Family background variables	Number of observations	R <sup>2</sup>
Seldom or never told bedtime stories (child's response)	0.233	(0.022)**	–	9056	0.03
	0.048	(0.036)	yes	9056	0.48
Often told bedtime stories (child's response)	-0.296	(0.022)**	–	9056	0.03
	-0.023	(0.040)	yes	9056	0.50
Seldom or never told bedtime stories (parent's response)	0.150	(0.019)**	–	9056	0.04
	0.051	(0.025)*	yes	9056	0.47
Often told bedtime stories (parent's response)	-0.271	(0.019)**	–	9056	0.02
	-0.029	(0.039)	yes	9056	0.52
Seldom went hiking with parents (child's response)	0.312	(0.021)**	–	9056	0.03
	0.012	(0.038)	yes	9056	0.57 <sub>0</sub>
Cognitive HOME index	-1.118	(0.051)**	–	9056	0.09
	-0.080	(0.070)	yes	9056	0.70 <sub>0</sub>
Emotional HOME index	-0.184	(0.049)**	–	9056	0.00
	0.070	(0.075)	yes	9056	0.61
There are few or no books at home	0.552	(0.024)**	–	9056	0.19
	0.235	(0.040)**	yes	9056	0.63
There is an Internet connection at home	-0.438	(0.013)**	–	9056	0.05
	-0.049	(0.027)	yes	9056	0.65

\* Significant at 5 %, \*\* Significant at 1 %.

Table A9. Detailed OLS regression estimates for *Table 8* (dependent variable: probability of attending a problematic class, independent variables: family background, home environment)

	Dependent variable: probability of being in a class highly segregated by ability			
	(1)	(2)	(3)	(4)
Roma	0.404 (0.028)**	0.143 (0.029)**	0.207 (0.029)**	0.123 (0.029)**
Biological mother in household		-0.074 (0.112)		-0.077 (0.114)
Non-biological mother in household		-0.061 (0.118)		-0.083 (0.121)
Biological father in household		0.066 (0.151)		0.096 (0.163)
Non-biological father in household		0.093 (0.151)		0.118 (0.163)
Mother's education: grades 0-8		0.166 (0.019)**		0.095 (0.021)**
Mother's education: vocational school		0.081 (0.014)**		0.036 (0.015)*
Mother's education: higher education		0.039 (0.010)**		0.018 (0.010)
Father's education: grades 0-8		0.095 (0.023)**		0.054 (0.023)*
Father's education: vocational school		0.042 (0.014)**		0.019 (0.014)
Father's education: high school diploma		0.002 (0.012)		-0.006 (0.012)
Mother employed		-0.016 (0.016)		-0.012 (0.016)
Father employed		-0.038 (0.019)*		-0.035 (0.018)
Proportion of years mother employed while child was age 0-14		-0.020 (0.021)		-0.017 (0.021)
Proportion of years father employed while child was age 0-14		-0.008 (0.027)		0.003 (0.026)
Logarithm of family income		-0.012 (0.011)		-0.008 (0.011)
Logarithm of household size		0.008 (0.025)		0.015 (0.025)
Number of unemployed adults		-0.001 (0.009)		0.003 (0.009)
Size of apartment, m2 per person		0.000 (0.001)		0.001 (0.001)
Number of rooms per person		-0.074 (0.024)**		-0.046 (0.024)
Bathroom in apartment		-0.105 (0.031)**		-0.077 (0.031)*
No money for food		0.022 (0.024)		0.011 (0.023)
No money for heating		0.029 (0.017)		0.021 (0.017)
Received regularized child-rearing assistance		-0.021 (0.015)		-0.031 (0.015)*
Free lunch in 8th grade		0.055		0.050

	(0.024)*	(0.024)*
Free textbooks in 8th grade	0.031 (0.012)**	0.032 (0.012)**
Mother's education - data missing	-0.024 (0.105)	-0.084 (0.108)
Father's education - data missing	0.053 (0.150)	0.070 (0.163)
Family income - data missing	-0.011 (0.014)	-0.008 (0.014)
Size of apartment - data missing	0.118 (0.056)*	0.100 (0.057)
Number of rooms - data missing	-0.093 (0.052)	-0.106 (0.055)
Bathroom - data missing	0.026 (0.095)	0.005 (0.095)
Poverty indicator - data missing	0.019 (0.055)	0.021 (0.056)
Region: Central	0.051 (0.030)	0.063 (0.030)*
Region: Central Transdanubia	0.011 (0.029)	0.024 (0.029)
Region: Western Transdanubia	-0.049 (0.027)	-0.029 (0.026)
Region: Southern Transdanubia	0.069 (0.035)*	0.074 (0.034)*
Region: Northern Hungary	0.107 (0.031)**	0.115 (0.031)**
Region: Northern Great Plain	0.101 (0.029)**	0.095 (0.029)**
Budapest	-0.101 (0.030)**	-0.081 (0.030)**
County seat	-0.084 (0.020)**	-0.068 (0.020)**
Other city	-0.041 (0.020)*	-0.033 (0.020)
Remote settlement	0.016 (0.023)	0.016 (0.023)
Seldom or never told bedtime stories (child's response)		0.043 (0.022)*
Often told bedtime stories (child's response)		0.003 (0.012)
Seldom or never told bedtime stories (parent's response)		0.045 (0.033)
Often told bedtime stories (parent's response)		-0.018 (0.011)
Seldom went hiking with parents (child's response)		0.039 (0.011)**
Cognitive HOME index		-0.052 (0.007)**
Emotional HOME index		0.006 (0.006)
Storytelling variable missing		-0.019 (0.022)
Cognitive HOME variable missing		0.080 (0.056)
Emotional HOME variable missing		-0.044 (0.029)

			(0.028)	(0.027)
Number of books less than 50			0.156	0.072
			(0.025)**	(0.026)**
Number of books around 50			0.089	0.041
			(0.020)**	(0.021)
Number of books: 50-150			0.064	0.037
			(0.015)**	(0.016)*
Number of books: 150-300			0.030	0.006
			(0.014)*	(0.015)
Number of books: 300-600			0.021	0.010
			(0.013)	(0.013)
Number of books: 600-1000			-0.006	-0.008
			(0.012)	(0.012)
Internet connection at home			-0.071	-0.028
			(0.012)**	(0.011)*
Number of books - data missing			-0.029	-0.064
			(0.048)	(0.044)
Internet connection - data missing			0.094	0.111
			(0.096)	(0.092)
Constant	0.177	0.421	0.154	0.275
	(0.008)**	(0.173)*	(0.017)**	(0.179)
Number of observations	9056	9056	9056	9056
R <sup>2</sup>	0.07	0.18	0.16	0.20

Robust standard errors clustered by school in parentheses

\*Significant at 5% level; \*\*Significant at 1% level

Figure A1. Distribution of Roma and non-Roma students by family background index  
Roma distribution: continuous line (average: 0.23);  
non-Roma distribution: dashed line (average: 0.57)

