GRINCOH Growth-Innovation-Competitiveness
Fostering Cohesion in Central and Eastern Europe

The achievement gap between Roma and non-Roma students in East Central Europe and its potential causes

Gábor Kertesi* and Gabor Kézdi**

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement "Growth-Innovatiion Competitiveness: Fostering Cohesion in Central and Eastern Europe" (GRINCOH)

Institution: * Institute of Economics of the Hungarian Academy of Sciences CERS, ** Central European University and Institute of Economics of the HAS CERS
WP 4, Task 3, P.4.11.
Final version

The achievement gap between Roma and non-Roma students in East Central Europe and its potential causes

Gábor Kertesi
(Institute of Economics of the Hungarian Academy of Sciences CERS)
kertesi@econ.core.hu

Gábor Kézdi
(Central European University and Institute of Economics of the HAS CERS)
kezdig@ceu.hu

Abstract

This study quantifies the achievement gap between Roma and non-Roma students in East Central Europe and assesses the potential causes of the gap. Using the UNDP survey of 2011, the only comparable data on the Roma spanning many countries, we show that the gap in the chances to get secondary education is substantial in all countries. When comparing young adults living with parents of comparable income and educational attainment, the gap drops by more than a half in most countries. Using unique data from Hungary, we assess the gap in standardized test scores and show that it is comparable to the size of the Black-White test score gap in the U.S.A. in the 1980's. The test score gap in Hungary is almost entirely explained by social differences in income, wealth and parental education, and ethnic factors do not play a significant role. We identify two major mechanisms by which the social disadvantages of Roma students lead to lower skills. Their home environment is less favorable for their cognitive development, and their schools are characterized by a lower quality educational environment. Ethnic differences in the home environment are, again, explained by social differences, and ethnicity seems to play no additional role. On the other hand, while access to higher quality schools is strongly related to social differences, Roma students seem to face additional disadvantages. The results suggest that besides policies that aim at alleviating poverty, well-designed interventions influencing the mechanisms can also improve the skill development of Roma and other disadvantaged children.

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 the Roma live in former communist countries in East Central Europe. The size of the Roma population is notoriously hard to assess because ethnic data collection is problematic.[1] One of the more reliable estimates of the size of the Roma population in East Central Europe put it slightly over 4 million in the early 1990s (Barany, 2002). According to these figures, the percentage of Roma in the total population was close to 10 percent in Bulgaria and Slovakia, between 4 and 7 percent in Hungary, Macedonia, Romania and Serbia, and around 2 percent in Albania and the Czech Republic. Representative evidence on the well-being of the Roma has been rare until recently (but see *UNDP* [2002], *Ringold et al.* [2005], *Higgins and Ivanov* [2006] and *Milcher* [2006]). The UNDP survey of 2011 provided the most up-to-date and most comprehensive evidence on the scale of the disadvantage of the Roma of East-Central Europe.

According to data of the UNDP survey of 2011, the employment rate among the Roma aged 20 to 64 was between 20 and 30 per cent in most East Central European countries. While labor market discrimination is likely to play a role (*FRA [2009]*), it is unlikely to explain such low levels. Numerous international studies also demonstrate that the basic skills acquired in early childhood and elementary school play a major role in shaping employment prospects. Education has a strong causal effect on earnings (*Card* [1999]); the skills gap can account for minorities' labor market disadvantages, at least in the US (*Neal–Johnson* [1996]); and skills also play a decisive role in other areas of life as well (*Heckman–Stixrud–Urzua* [2006]). Skills presumably play a similarly significant role in determining children's life chances in East Central Europe. Using multiple datasets in Hungary, *Kertesi-Kézdi* [2010] decomposed the employment gap between Roma and non-Roma in Hungary and found that the larger part of the gap is explained by educational differences. Although no direct proof is available at present for the role of skills in the ethnic employment gap, it is likely to be significant in East Central Europe. Understanding the extent and the origins of the gap in skills is therefore a very important step towards understanding the origins of the gap between the Roma minority and the non-Roma majority in many areas of life.

This study quantifies the achievement gap between Roma and non-Roma students in East Central Europe and aims to explain it by policy-relevant factors. It focuses on two major questions: Is the achievement gap between Roma and non-Roma students due to ethnic specificities of the Roma or is it the result of social disadvantage? And what are the mechanisms behind the emergence of the

[1] Unfortunately, census data are not well suited for estimating the size of the Roma population. In most countries the national census collects data on “nationality” by asking respondents to report their national or ethnic identity. Recent research on the Hungarian Roma (Simonovits and Kezdi, 2013) shows that the Roma in Hungary have multiple identities: 99 per cent of the respondents who consider themselves Roma considered themselves Hungarian as well. The national censuses in the East Central European countries do not allow for stating multiple identities but force respondents to choose one. The number of respondents whose answers indicate Roma identity in the national censuses is therefore significantly smaller than the number of respondents who have Roma identity.

gap?

When answering the first question, one has to remember that many things may appear to be ethnic characteristics, which are in fact nothing of the sort. The existence of a Roma/non-Roma school achievement gap frequently leads public opinion and those working in the education system and social policy to seek an explanation in characteristic ethnic behavior patterns, a Roma mentality, the "characteristic value system" of the Roma, etc. If this diagnosis were correct, then public policy programs aiming to reduce the gap should primarily strive to influence such characteristic ethnic behavior patterns, to "shape attitudes" and transform the "Roma mentality". However, as we will show, the large achievement gap between Roma and non-Roma students can be explained almost entirely by well-defined social differences, not ethnicity. Poverty, low levels of parental education, parents' weak attachment to the labor market, and associated further disadvantages largely account for the skills deficits of Roma students. Thus, references to so-called "characteristic ethnic behavior patterns" consist of false diagnosis; similarly, interventions that aim to transform the "characteristic mentality" are likely to be bad therapies.

Answering the second question is equally important*:* uncovering that disadvantaged family background is responsible for the skills deficits does not provide a full explanation. We need to understand the mechanisms by which these social and familial disadvantages affect skill development because an understanding of these mechanisms is indispensable for designing effective social policy programs to reduce the skills deficits. This knowledge can help us determine what kind of resources the affected families should be provided with, how to help them understand the importance of these resources, and how to help them acquire the knowledge and skills to use these resources.

This paper identifies the major mechanisms that are likely responsible for the large part of the achievement gap between Roma and non-Roma students. After a short introduction to the history of the Roma of East Central Europe, we show that the achievement gap – measured by secondary school completion rates – is huge in all East Central European countries. We also show evidence suggesting that the gap is in large part due to social factors as opposed to ethnic characteristics. Then we focus on the gap in standardized test in Hungary where we have substantially more detailed data. The test score gap is substantial (similarly to the Blac-White gap in the U.S. in the early 1980's), it is almost entirely explained by social differences in income, wealth and parental education, and ethnic factors do not play a significant role. We identify two major mechanisms by which the social disadvantages of Roma students lead to lower skills: home environment and schools. We show home environment and parenting practices explain one third to two thirds of the test score gap. We also show that the gap between Roma and non-Roma students in the same school

and same classroom is 60 per cent less than the national gap. Ethnic differences in the home environment are fully explained by social differences, and ethnicity seems to play no additional role. On the other hand, while access to higher quality schools is strongly related to social differences, Roma students seem to face additional disadvantages.

Some background on the Roma of East Central Europe

The Roma have no historical homeland in Europe. They originated in India and migrated to Eastern Europe 700 years ago. The Roma are a heterogeneous people spread across many countries. Some speak dialects of the Romani language, whereas others adopted the language of their host country, often in the form of a special dialect. The vast majority of the Roma of Central and Eastern Europe settled a long time ago, and their romanticized image as travellers is based on exceptions, which are often cases from Western Europe. The Roma were enslaved in some parts of Central and Eastern Europe for centuries, and they were often targeted by law enforcement. Historical evidence on the well-being of the Roma communities and their relationship to mainstream societies is relatively scarce. The following two paragraphs describe a widely accepted but not uncontested view of their history (see, for example, Barany, 2002, Hancock, 2002, Janky and Kemény, 2003, and Kemény, 2005).

For centuries, the integration and assimilation of the Roma remained limited. In many respects, they lived outside mainstream society both before and well into the Industrial Revolution. The Roma had no land or any other formal property, and when they were not slaves, they worked as independent laborers or sold their own products and services. The Industrial Revolution and the emergence of centralized nation states brought the Roma minority closer to mainstream society, but they also undermined their traditional communities. During the Second World War, the Roma were subjects of deportations and mass executions, similar to the Jews (the Roma Holocaust is known as Porajmos). The communist regimes hastened the dissolution of the Roma communities and instigated a paternalistic assimilation process. Many Roma faced relocation into villages and towns inhabited by the majority (often into segregated settlements), obligatory employment in the state sector, and compulsory schooling for their children. As a result, many (in some countries, most) Roma families have had stable wage earners under the communist regimes and have seen their children achieve literacy or vocational degrees. At the same time, many of the ties within the Roma communities have been destroyed.

The fall of the communist system led to a deep recession and a thorough transformation of labor demand in most transition countries. Demand for unskilled labor collapsed. The more successful

post-communist economies started to grow quickly during the mid-1990s, but even they did not experience an increase in demand for unskilled labor. Many unskilled people who lost their employment during the transition period have been left without a regular formal job ever since. A widely accepted view is that the dramatic drop in demand for low-skilled workers affected the Roma especially severely.

The achievement gap in East Central Europe

Due to data constraints, the achievement gap between Roma and non-Roma students has been impossible to assess until recently. Data from the 2011 survey of UNDP provide a unique opportunity to estimate the achievement gap between young Roma and non-Roma people. The survey collected information in 12 countries of East and Central Europe in a fully standardized way. In each country, the sample consists of people living in areas with high concentration of Roma. Because of this sampling design, we can expect that the Roma covered by the survey are more disadvantaged than the entire population with Roma identity. Similarly, the non-Roma population covered by the survey is likely to be poorer and more disadvantaged than the entire population. When interpreting the findings from the UNDP 2011 data one has to keep in mind these issues with representativeness.

Our focus is on the achievement between young Roma and non-Roma. The data do not contain standardized test scores but we have information on educational attaintment. We can therefore analyze the gap between school completion rates in younger cohorts. The gap measured in the UNDP survey may be larger or smaller than the gap in the entire population, depending on the pecularities of the local sampling procedure. For robustness checks, therefore, we show achievement figures for the entire population of the relevant age groups. The source of the data on the general population is the Barro-Lee dataset (*Barro-Lee [forthcoming]*).

Table 1 shows the fraction of the 20 to 24 years old respondents with upper-secondary education by country. The first two columns show the fraction with upper secondary education in the UNDP dataset among the Roma and the non-Roma, respectively. The third column shows the national fraction with secondary education from the Barro-Lee data.

The national averages may be different from the average of the Roma and non-Roma figures for at least two reasons. First, both the Roma and the non-Roma subsamples in the UNDP cover the part of the population that lives in or near areas with high Roma concentration. As a result, both the Roma and the non-Roma figures may be different from their respective national values, and they are most likely lower than that. Second, the size of the non-Roma subsample in the relevant age range

in the UNDP data is rather small (below 100 in most countries) that induces considerable random variation in the estimates (the 95% confidence interval can be as wide as ±20 percentage points). The Barro-Lee figures are based on data from national censuses and estimates from very large surveys (such as the micro-census held in most of the countries listed below) and thus their confidence intervals are tiny.

Table 1.
The percentage of 20 to 24 years old with upper secondary education. Roma and non-Roma respondents in the UNDP 2011 survey and population figures.

	Percentage of 20 to 24 years old with secondary education [a]			The ethnic gap	
	UNDP survey, 2011		National	Non-Roma	National avg.
Country	Roma	Non-Roma	average	vs. Roma	vs. Roma
	(1)	(2)	(3)	(2) – (1)	(3) – (1)
Albania	3.1	43.5	22.6	40.4	19.5
Bosnia and Herzegovina	15.8	86.2	n.a.	70.4	n.a.
Bulgaria	20.7	65.2	42.3	44.5	21.6
Czech Republic	30.1	76.1	79.0	46.0	49.0
Slovakia	18.1	48.0	42.5	30.0	24.4
Montenegro	7.0	79.3	n.a.	72.3	n.a.
Croatia	19.7	72.8	42.0	53.1	22.4
Hungary	21.3	62.0	70.4	40.7	49.2
Macedonia	19.1	82.5	n.a.	63.4	n.a.
Moldova	7.5	56.9	n.a.	49.4	n.a.
Romania	11.5	67.1	42.2	55.7	30.7
Serbia	12.9	82.8	41.2	69.9	28.3

[a] Upper secondary education (ISCED level 2) or vocational education .

The figures show large gaps in all countries. According to the UNDP Survey figures, the gap between Non-Roma and Roma people of 20 to 24 years of age ranges from 30 percentage points (Slovakia) to 72 percentage points (Montenegro). As we indicated above, the size of the sample (especially the non-Roma subsample) is small in most countries. As a result, the confidence intervals around these differences range between ±10 percentage points (Montenegro and Serbia) to ±15 percentage points (most other countries). Sampling uncertainties notwithstanding the systematically large estimates in all countries point to a substantial achievement gap. The (unweighted) average gap in the 12 countries combined is over 50 per cent (confidence interval ±4 percentage points).

Comparing the Roma figures to national averages show smaller but still substantial gaps. While these gaps should be somewhat smaller because the national figures contain the Roma achievements as well, the differences relative to the gaps measured in the UNDP sample are too large to be

explained by this. The specificities of the UNDP sample may also play a role in the differences. But the sampling uncertainties in the UNDP estimates can alone explain most of the differences.

Albeit in a limited way, the data help answering our first question: whether the gaps documented in Table 1 are due to ethnic factors or social disadvantage. In order to answer this question one would have to use data on social disadvantage during childhood. The UNDP survey is rich in information on poverty, but it is a cross-sectional survey without retrospective information on conditions in childhood. As a result, we need to analyze current family conditions of the 20 to 24 years old instead of their conditions in childhood. For the 20 to 24 years old living with their parents the current family conditions may be a good representation of the family conditions in childhood because poverty and social disadvantage are likely to be persistent. However, we cannot carry out the analysis for those who don't live with their parents As Table 2 shows, approximately half of the respondents live with their parents.

Table 2.
The percentage of 20 to 24 years old with upper secondary education. Roma and non-Roma respondents by whether they live with their parents or not.

	Living with parents			Not living with parents			Per cent
Country	Roma (1)	Non-Roma (2)	Gap (3) = (2)-(1)	Roma (4)	Non-Roma (5)	Gap (6) = (5)-(4)	living with parents (7)
Albania	3.6	46.7	43.0	2.5	35.3	32.8	58.9
Bosnia and Herzegovina	22.7	90.0	67.3	8.0	75.0	67.0	58.2
Bulgaria	28.1	65.5	37.4	15.2	64.7	49.5	45.3
Czech Republic	31.2	85.7	54.5	29.0	69.8	40.8	45.8
Slovakia	29.2	50.9	21.7	12.5	44.2	31.7	45.0
Montenegro	11.3	85.7	74.4	4.8	71.7	66.9	39.1
Croatia	27.7	72.7	45.0	13.9	73.1	59.2	48.8
Hungary	29.4	75.0	45.6	13.9	40.7	26.9	50.9
Macedonia	24.3	91.5	67.2	12.4	57.7	45.3	60.0
Moldova	7.8	70.0	62.2	7.2	36.0	28.8	55.7
Romania	15.1	75.7	60.5	7.8	57.6	49.7	50.4
Serbia	19.6	87.8	68.2	6.7	63.2	56.4	55.0

[a] Upper secondary education (ISCED level 2) or vocational education .

The fraction of 20 to 24 year olds living with their parents varies considerably across countries, from less than 40 per cent in Montenegro to 60 per cent in Macedonia. In the larger countries, the fraction is relatively stable between 45 per cent and 55 per cent. The characteristics of those living with their parents and those not living with their parents differ considerably in most countries.

In order to decompose the achievement gaps to social background and ethnic components, we

estimated regressions for each country separately with secondary education on the left-hand-side and the Roma dummy variable on the right-hand-side together with a measure of income and parental education. These are linear probability models: the left-hand-side variable is whether the 20 to 24 years old respondent has an upper secondary or vocational degree. For each country, we estimated three versions of the regression. The first version includes the Roma dummy on the right-hand-side only. The coefficient on that variable reproduces the "raw" gaps in upper secondary education among the 20 to 24 years old living with their parents as documented in Table 2. The second regression includes a summary measure of income and poverty besides the Roma dummy variable. This summary measure was created using (log) monthly income per equivalent household member, (log) expenditure per equivalent household member, ownership of various items such as car, washing machine etc., and poverty indicators.[2] The third regression includes parental education besides the Roma dummy and the income measure.

Table 3 shows the summary statistics of the right-hand-side variables for the sample analyzed here and Table 4 summarizes the results of the regressions.

Table 3.
Mean values of the stadardized index of income.
Roma and non-Roma people 20 to 24 years of age living with their parents.

Country	Standardized index of income		
	Roma (1)	Non-Roma (2)	Gap (2) – (1)
Albania	-0.32	0.68	1.00
Bosnia and Herzegovina	-0.45	0.99	1.44
Bulgaria	-0.27	1.14	1.42
Czech Republic	-0.42	1.38	1.80
Slovakia	-0.48	0.30	0.77
Montenegro	-0.50	0.83	1.33
Croatia	-0.38	0.75	1.13
Hungary	-0.35	1.00	1.35
Macedonia	-0.39	1.01	1.40
Moldova	-0.29	1.19	1.47
Romania	-0.27	1.13	1.40
Serbia	-0.47	0.95	1.42

Data: UNDP Roma Survey, 2011; 20 to 24 years old respondents living with their parents.

The standardized index of income and poverty shows enormous differences between Roma and

[2] The income measure is a linear combination of the respective items. It was created by regressing the Roma dummy on the items and the country dummies using a pooled probit regression and predicting the probit index (i.e., the items were combined by their respective probit coefficients). The index was then standardized for each country separately. Created this way, the income measure summarizes the income status of families in a way that creates the maximum difference between the Roma and the non-Roma respondents in the UNDP survey, and its unit of measurement is standardized for each country separately.

non-Roma. Except for Slovakia, the difference is one standard deviation or more. IN the Czech Republic, it is almost 2 standard deviations. Whether these cross-country differences are due to the differential gap between the income status of the Roma minority and the non-Roma majority or due to differences in sampling is impossible to tell without external validation. But the magnitude of the gaps in all countries implies very large disadvantages among the Roma compared to the mainstream society.

Table 4.
The gap in the probability of upper secondary education[a] degree between Roma and non-Roma people 20 to 24 years of age living with their parents. Estimates of the coeficient on the Roma dummy from three regressions estimated for each country.

	The gap in the probability of upper secondary education between Roma and non-Roma of age 20 to 24 living with their parents		
Country	Raw gap (1)	Gap conditional on household income (2)	Gap conditional on household income and parental education (3)
Albania	-0.43	-0.30	-0.20
S.E.	(0.06)**	(0.06)**	(0.06)**
Bosnia and Herzegovina	-0.68	-0.43	-0.35
S.E.	(0.06)**	(0.09)**	(0.11)**
Bulgaria	-0.37	-0.10	-0.02
S.E.	(0.10)**	(0.12)	(0.11)
Czech Republic	-0.54	-0.15	-0.14
S.E.	(0.08)**	(0.15)	(0.17)
Slovakia	-0.22	-0.03	0.03
S.E.	(0.08)*	(0.09)	(0.08)
Montenegro	-0.74	-0.64	-0.36
S.E.	(0.06)**	(0.09)**	(0.12)**
Croatia	-0.47	-0.29	-0.17
S.E.	(0.08)**	(0.10)**	(0.11)
Hungary	-0.45	-0.23	-0.13
S.E.	(0.09)**	(0.11)*	(0.10)
Macedonia	-0.67	-0.51	-0.30
S.E.	(0.05)**	(0.08)**	(0.08)**
Moldova	-0.62	-0.44	-0.28
S.E.	(0.08)**	(0.11)**	(0.11)*
Romania	-0.61	-0.46	-0.39
S.E.	(0.07)**	(0.10)**	(0.11)**
Serbia	-0.68	-0.42	-0.33
S.E.	(0.05)**	(0.08)**	(0.09)**

[a] Upper secondary education (ISCED level 2) or vocational education .
Data: UNDP Roma Survey, 2011; 20 to 24 years old respondents living with their parents.
Standard error estimates in parentheses are robust to heteroskedasticity and within-household correlation for siblings.
* Significant at 5%; ** Significant at 1%.

According to the results, 15 to 85 per cent of the achievement gap is explained by lower incomes among the Roma, and 35 per cent to 100 per cent is explained by lower incomes and lower parental education. 50 per cent of more of the achievement gap is explained by the income measure in Bulgaria, the Czech Republic, Slovakia and Hungary, around 40 per cent is explained in Bosnia, Croatia and Serbia, and at most 30 per cent is explained in Albania, Montenegro, Macedonia, Moldova and Romania. When comparing people with the same parental education as well as the same income, the achievement gap between Roma and non-Roma 20 to 24 year old respondents is less than 10 percentage points and statistically insignificant in Bulgaria and Slovakia, and it is between 10 and 20 percentage points but also statistically insignificant in the Czech Republic, Croatia and Hungary.

These results suggest that the larger part, and perhaps all, of the gap between young Roma and non-Roma in terms of their secondary school completion rates can be explained by non-ethnic family background variables, and family income status plays a major role.

At the same time, the analysis cannot answer the question about the mechanisms that make children in poor families achieve lower levels of education. The income measure we used is also imperfect in the sense that it corresponds to family incomes when the children were in their twenties, whereas the appropriate question would involve income status during childhood. Moreover, the analysis was be carried out for a subpopulation of those still living with their parent. Fortunately, unique data from Hungary allows us to do a more thorough analysis of the achievement gap between Roma and non-Roma students.

The test score gap in Hungary

Standardized competence test scores and a survey with ethnic identifiers linked to the test scores data provide a unique opportunity to analyze the test score gap between Roma and non-Roma students in Hungary. The source of the test score data is the May 2006 National Assessment of Basic Competences (NABC), administered to every 8th grade elementary school student. The survey linked to those test scores is the Hungarian Life Course Survey of Tárki.

The HLCS followed 10 000 young people with yearly regularity, starting in the fall of 2006. The bulk of the survey was composed of students who were in the 8th grade in May 2006 and who participated in the NABC, filling out the reading comprehension and mathematics tests as well as the family background questionnaire. The HLCS also included special needs students who did not participate in the NABC, but did fill out a simplified version of the reading comprehension test. To be included in the sample, students had to fill out a family background questionnaire and their

parents had to provide a written agreement to participate in the survey. As one of the most important purposes of the survey was to analyze school disadvantage, students with lower test results and special needs students are overrepresented in the sample. We weighted the samples to account for the consequences of unequal selection rates.

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

In our analysis, we relied on the data collected in the first two waves of the survey. We limited 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 restrictions on the sample were made necessary by our method of identifying the Roma ethnic identity. The parents were asked what nationality or ethnicity they identified with primarily or secondarily in both waves of the survey. These two questions made it possible for them to choose a double 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 the 2007 survey. Using this definition, Roma youth comprise nearly 8 % of all 8th grade students; the size of the Roma subsample is 848 students (see Table A2 of the Appendix). The total sample size is 9056 students with reading comprehension test results, and 8335 students with mathematics test results. The difference is due to the fact that the special needs students only completed the reading comprehension test.[3] Table A3 of the appendix shows the magnitude of the distortion arising from sample selection and the basic data of the students who were eliminated from the sample for various reasons.

The test scores measure skills that have a large impact on their choice of secondary school and on key events in their secondary school careers, as *Figure 1* demonstrates. The figure shows the likelihoods of earning different types of secondary school degrees by age 21 as a function of the 8th grade test scores. The vertical axis shows the fraction of respondents with general high school degree, technical high school degree (these two degrees involve passing a maturity examination that is also the entry test for college), vocational school degree, and the fraction of respondents without any secondary degree. The horizontal axis shows 10 equal-sized categories created by the average of the mathematics and reading test scores measured in 8th grade so that group 1 has the lowest scores and group 10 has the highest scores.

3 6 % of all 8th graders (and 12 % of Roma 8th graders) were special needs students in 2006; the majority had a mild intellectual disability. Special needs students generally do not participate in the NABC, but a special reading comprehension test was designed for them in 2006. The results of this test can be assigned a place in the distribution of the general reading comprehension test.

Figure 1

The likelihood of acquiring different types of secondary school degree by age 21 as a function of 8th grade reading comprehension test scores*

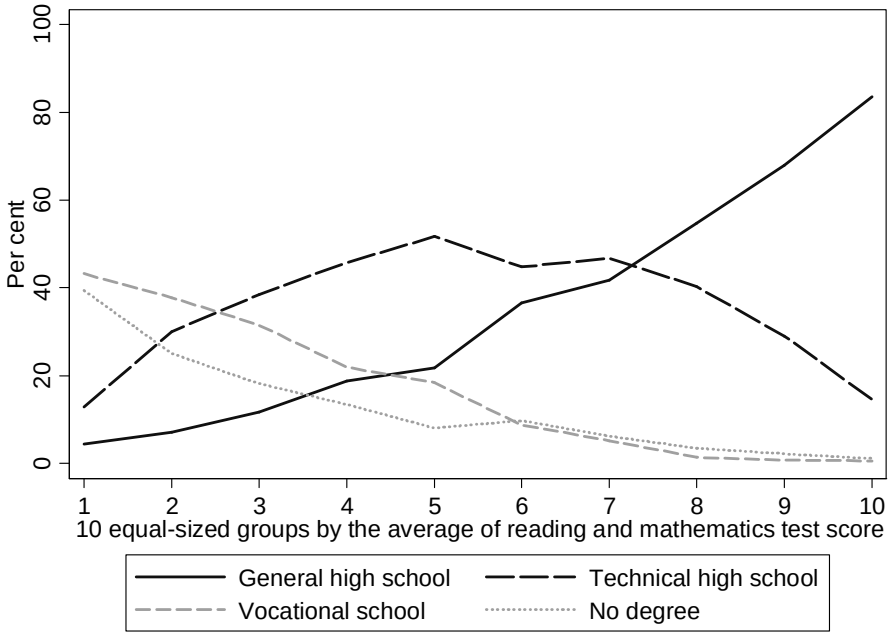


Figure 1 reveals a strong mechanism of skill-based selection at work in secondary school degrees. Over 80 per cent of the students with the strongest test scores complete general high school, and the rest complete technical high school. At the other extreme, 40 per cent of students with the weakest test scores do not have a secondary degree by age 21, another more than 40 per cent complete vocational training school, and the remaining few earn a degree in technical high schools. The relationship between the likelihood of being without 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 genera high school degree is strongly positively related to the test scores. The pattern with respect to general high school follows an inverted U-shape with the highest likelihood for students in the middle of the test score distribution. *Figure* 1 implies that selection into secondary school types and subsequent success is strongly related to skills in 8th grade, and the test scores of the NABC are good measures of those skills.

Having established the importance of the test score measures, we turn to ethnic differences in the test scores. *Table 5* shows the magnitude of the standardized test score gap between Roma and non-Roma students using the data on 8th graders from the 2006 NABC. As a comparison, we provide similar data on the test score gap between 13-year old and 8th grade black and white students in the

United States. We include the test scores of 13-year-old students from the US because this is the format of the data concerning the less recent past, at the turn of the 1980s.

Table 5
The magnitude of the Roma/non-Roma and black/white test score gaps in Hungary and the United States (measured in standard deviation units of the national average of the given test)

Year	Roma / non-Roma gap, 8th grade, Hungary [a]		black / white gap, 8th grade, USA[b]		black / white gap, 13-year-olds, USA[c]	
	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

[a] Calculated by the authors. *Source:* the combined data of the 2006 NABC and the HCLS.
[b] National Assessment of Educational Progress (NAEP), Main NAEP tables, 1992 and 2007.
[c] National Assessment of Educational Progress (NAEP), Long-Term Trend tables, reading: 1980, 1992 and 2008, mathematics: 1978, 1992 and 2008.

The raw difference between Roma and non-Roma students' scores is approximately one unit of standard deviation.[4] This matches the size of the gap between black and white 13-year-old students in the United States approximately 30 years ago. However, that gap has narrowed significantly since the early 1980s. The ethnic differences measured at the end of the 2000s in the US are approximately 20 percentage points smaller than ethnic differences in Hungary.

While our data allows for analyzing the test score gap in one year and for a single age group, we can shed some light on the age pattern of the gap using other, less comprehensive data. Two other longitudinal studies with an adequately large sample size measured skills and used a technique similar to the HLCS's to determine the students' ethnic identity. Importantly, both the samples and the skills tests are widely dissimilar. As a result, meaningful comparisons across age groups can only be made within a single sample.

The first data come from the impact assessment of the National Education Integration Network (*Kézdi-Surányi* [2008]). This data enables us to compare 2nd and 4th grade students. The study measured the arithmetic and reading skills of approximately 4000 students in 60 treatment and control schools in two waves (spring 2005 and spring 2007).[5] The second data allow for comparing 6th and 8th grade students; these data are based on the "Inter-Ethnic Relations, 2010" survey by Educatio. The survey collected data on the networks of 8th grade students at 88 schools in 74 cities, and respondents were linked to their administrative files with their 6th grade test scores from 2008 and the 8th grade test scores from 2010. The third dataset is allows for comparing 8th and 10th graders: this is part of the HLCS linked to the NABC data used in our main analysis. Unique student identifiers that allow for connecting their test scores measured at different ages were

4 The reading score gap is somewhat smaller than one unit of standard deviation and the mathematics score gap is somewhat larger. These values are provided based on the standard deviation of the full sample.
5 *Kézdi–Surányi* [2008] provides a detailed account.

introduced later for younger cohorts so that the complete linking of 8th and 10th grade test scores is not possible for this sample. However, the data collecting agency was successful to link 58 % of the 8th graders who took part in the 2006 survey to their 10th grade test scores in 2008.

We summarize the results of all the measurements in *Table 6*. In addition to the raw test score gap, we include the values of the gap after correcting for gender, age, household presence and education of the mother/father in parentheses.

Table 6
The Roma/non-Roma test score gap by grade level (measured in standard deviation units of the national average of the given test). First data: raw gap; data in parentheses: includes controls for gender, age, no mother/father, parental education

Survey / Year	grade	test		
		SZTE arithmetic[d]	SZTE reading[d]	NABC reading[e]
OOIH[a]	2th	–0.76 (–0.49)	-	-
2005/2007	4th	-	–0.86 (–0.53)	-
IEK-OKM[b]	6th	-	-	–0.67 (–0.33)
2008/2010	8th	-	-	–0.68 (–0.35)
HLCS-OKM[c]	8th	-	-	–0.82 (–0.22)
2006/2008	10th	-	-	–1.01 (–0.33)

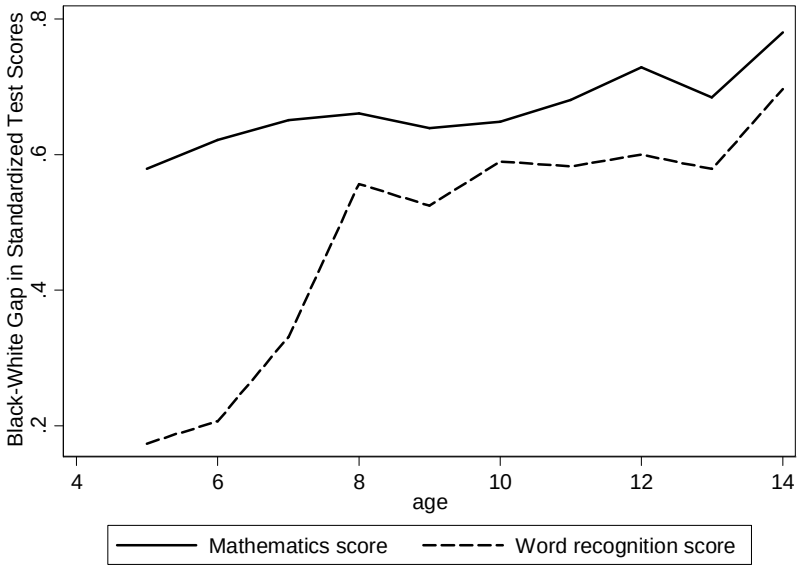
[a] The impact analysis of the National Education Integration Network (OOIH); sample: students in 2nd grade in spring 2005 and 4th grade in spring 2007. See *Kézdi–Surányi* [2008].
[b] The sample of the Educatio "Inter-Ethnic Relations, 2010" (IEK) survey combined with the 2008 National Assessment of Basic Competencies (NABC) 6th grade and the 2010 NABC 8th grade test score data.
[c] The sample of the Tárki Hungarian Life Course Survey (HLCS) combined with the 2006 NABC 8th grade and 2008 NABC 10th grade test score data. The table includes the data of students from the HLCS only if they could be identified as 10th graders in the 2008 NABC.
[d] Reading comprehension test for 2nd graders and arithmetic skills test for 4th 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: 2nd graders, 2006: 4th graders. (See: *Csapó* [2007].)
[e] NABC reading comprehension and mathematics tests.

The available data point to the relative stability of the test score gaps measured in grades 5-8: the ethnic differences measured in the 6th and 8th grades are almost exactly identical. By contrast, the gap increases in the first years of elementary school (between 2nd and 4th grade), and in the first two years of secondary school (between 8th and 10th grade).[6] The significant divergence between the raw gaps and the values corrected for a few important socio-demographic variables suggests that ethnicity is unlikely to be the cause of the differences.

There is no way to know how the test score gap between Roma and non-Roma students would change across time in a study (optimally with a nationally representative sample) started in early childhood, with surveys administered regularly throughout the children's school career. However, the international literature suggests that the deficits of disadvantaged ethnic minorities generally tend to increase with age. We present the results of such a study from the United States in *Figure 2*.

6 As the gap between social groups is generally larger in arithmetic and mathematics tests than reading comprehension tests, we find it likely that the gap would also increase between grades 2 and 4 using identical tests.

Figure 2
The mathematics and reading comprehension test score gap between black and white students from age 5 to 14*



* PIAT math and word recognition scores (in % of the standard deviation of the full sample), panel data from the Children of the National Longitudinal Survey of Youth 1979 (CNLSY79).
Source: Fryer [2011], raw data in Table 11, row 1, p. 895–896.

Hungarian research is in line with international surveys (*Lee–Burkam* [2002]) in finding that the children of Roma parents struggle with significant deficits by the time they reach kindergarten age. The impact analysis of the Biztos Kezdet (Sure Start) program collected data on nearly 1000 4- to 6-year-old kindergarteners in municipalities which had a Sure Start Children's Center and in a control group of similar municipalities without such a program in place.[7] Due to the composition of the program's target group, the social status of the children in the combined sample is somewhat lower than in a representative sample. However, its social heterogeneity allows us to illustrate the magnitude of the skills deficits that Roma children have at a very early age. The Sure Start study measured the vocabulary of kindergarten-age children. The study design used the LAPP-test developed by József Lőrik and his colleagues *(Lőrik* [1999]) to examine children's picture naming performance. The Roma children's raw deficit compared to the non-Roma children was 66 % of the standard deviation of the full sample, which shrank to 11 % after correcting for gender, age, household presence and education of the mother/father.

We can observe that correcting for a few simple demographic and family background variables reduced the raw test score gap significantly in every sample, without respect to age (kindergarten age through 10th grade of secondary school). Differences in the composition of the sample clearly

7 See *Sure Start* [2010]. We would like to thank *Éva Surányi*, the leader of the study, for giving us access to the data files used in the analysis.

play an important role in the Roma/non-Roma test score gap.

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 make up a much larger percentage of the students (and thus, of the sample), we will 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 bad circumstances as Roma students?

For our calculations, we used the family background variables presented in Table 7. Together, the variables represent the family's long-term income/wealth, its life chances in a broad sense.

Table 7
Family background variables

Name of variable	Definition of variable
Biological mother in household	Lives with biological mother: yes/no
Non-biological mother in household	Lives with non-biological mother: yes/no
Biological father in household	Lives with biological father: yes/no
Non-biological father in household	Lives with non-biological father: yes/no
Mother's education	Mother's (biological/non-biological) highest completed level of educational attainment: 0-8 years of elementary school / vocational school / high school diploma / higher education
Father's education	Father's (biological/non-biological) highest completed level of educational attainment: 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: ratio of years worked a while child was 0-14 years old, %
Father's long-term employment	Father: ratio of years worked a while 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^2	Surface area of apartment / number of household members, m^2/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)	Has it happened that there was not enough money for food in the past 12 months? yes/no
Poverty2 (income does not cover heating)	Has it happened that there was not enough money for heating in the past 12 months? yes/no
Poverty3 (regularized child-rearing assistance)	The family receives regularized child-rearing 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 in Hungary: Central Hungary / Central Transdanubia / Western Transdanubia / Southern Transdanubia / Northern Hungary / Northern Great Plain
Place of residence: settlement type	Settlement types: Budapest / county seat / other city / village
Place of residence: remote settlement	Settlement is difficult to access in terms of transportation[a]: yes/no

[a] See *Köllő* [1997].

We summarize the results of two different estimates in *Table 8*. 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 the propensity score matching is more flexible by allowing for nonlinearity of effects and ensure that

only comparable Roma and non-Roma students are compared (the overlap assumption). We have estimated two kinds of matching models, nearest neighbor matching and stratified matching.[8]

Despite the methodological differences, all estimates agree that the Roma students' school deficits are to a large extent explained their adverse long-term socio-economic circumstances.[9] The raw achievement gap is approximately one unit of standard deviation in magnitude; three-fourths of the mathematics and four-fifths of the reading gap would disappear if we assumed that the non-Roma majority students lived in similarly poor circumstances as the Roma students.

Table 8
The magnitude of the ethnic test score gap after correcting for differences in social composition

	Roma parameter (standard error)[a]	Number of observations[b]	R^2
	Reading comprehension		
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	–
	Mathematics		
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	–

[a] Standard errors in parentheses.
[b] 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.

Reversing the thought experiment, would the ethnic test score gap would exhibit a significant decrease if Roma students lived in similarly good circumstances as non-Roma students? As there are very few, if any, Roma students in the sample in circumstances similar to the better-circumstance non-Roma students, the OLS and matching results cannot be used to answer that question. But we may be able to make the appropriate comparisons in the middle ranges.

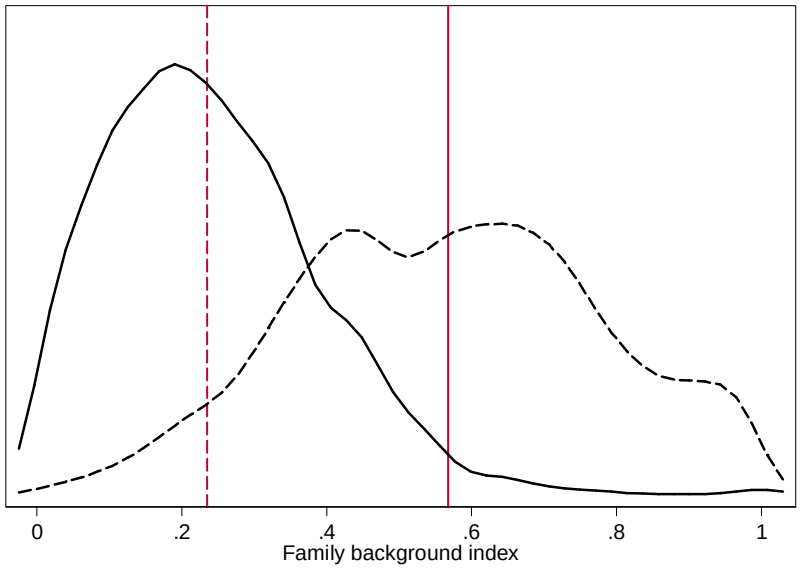
To see the test score gap along the distribution of family background, we combined the family background variables into a one-dimensional synthetic variable. We created the linear combination

8 On the methodology, see *Dehejia–Wahba* [2002].
9 The basic data for the family background variables by Roma/non-Roma students can be found in Table A3 of the appendix.

of family background variables by regressing test scores (the average of the reading and mathematics scores) on the family background variables and using the coefficients of that regression. Then we normalized the resulting values for a range of 0 to 1. People living in worse socio-economic circumstances are thus in the range closer to 0, while people living in better circumstances are in the range nearer to 1. Using the family background index thus defined, we can provide a visual representation of the size of the test score gap as a function of family background. The comparison is only meaningful for the range of values with an adequate sample size of both Roma and non-Roma students. *Figure 3* shows the results. The overwhelming majority of Roma students live in worse circumstances than the non-Roma student average: the Roma subsample is too small to be meaningful over values of 0.6.

Figure 3

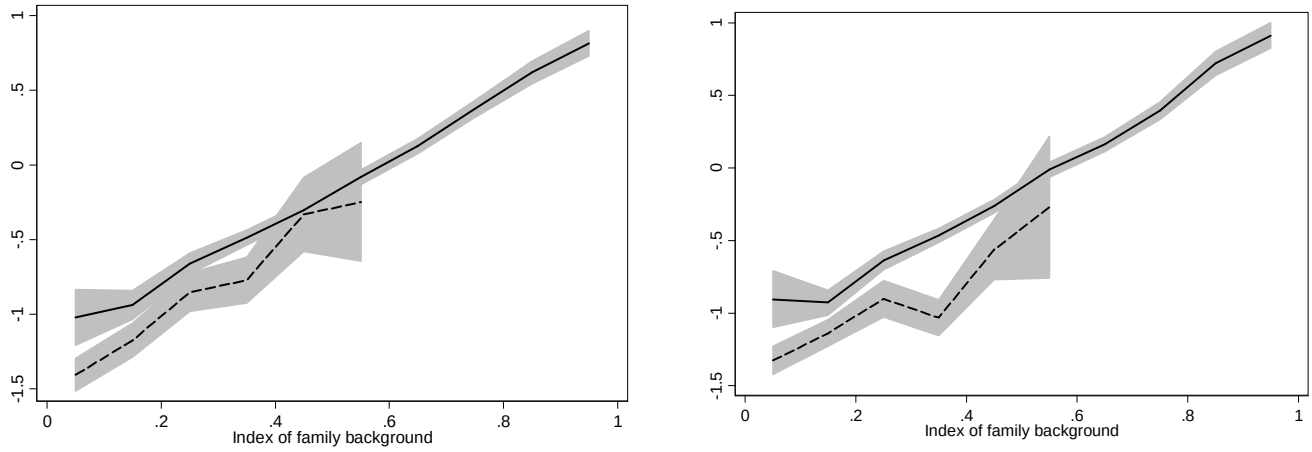
Distribution of Roma and non-Roma students by family background index
Roma distribution: continuous line (average: 0.23);
non-Roma distribution: dashed line (average: 0.57)



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 individually for each interval. We restricted the estimates for the Roma students to the 0–0.6 range. The estimates are presented in *Figure 4*. The gray zone marks the 95 % confidence intervals (within ± 2 standard errors of the mean).

Figure 4
Reading and mathematics test results as a function of the family background index
(The gray zone marks the 95% confidence intervals.)
Continuous lines: Non-Roma. Dashed lines: Roma

Reading **Mathematics**



Although our method would allow for the non-linear relationships, the Roma and non-Roma graphs can both be considered practically linear. They are also very close to each other. In the case of the reading score, the difference is very small and has a decreasing tendency; in the case of the mathematics score, the difference is somewhat larger and it is hard to tell whether the two lines converge or diverge. These results suggest that the test scores of Roma students from more affluent families (those in the middle third of the overall distribution of the family index) have very similar test score than similar non-Roma students. Nothing in these results suggests that the difference would be larger at higher parts of the distribution. We can speculate therefore that if Roma families lived in improved circumstances, then the reading and the mathematics test score gaps would narrow, perhaps to be insignificant. Thus, in a hypothetical case of full integration, when the family index distribution of the Roma students would be identical to the family index distribution of the non-Roma students, the test score between Roma and non-Roma students would be just as small as the estimates based on OLS and propensity score matching (see Table 8) had predicted.

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, and ethnicity seems to play a very small role at most. We now turn to answering our second question: What mechanisms are responsible for transforming the adverse long-term socio-economic circumstances of the Roma population into school achievement problems?

The transmission mechanisms

The literature identifies three major sets of mechanisms that lead to low achievement of disadvantaged students. All other factors being equal, children's skill accumulation and school

performance is weaker if (1) their health is worse than average; (2) they have little access to resources and activities important to developing their skills in their home learning environment; and (3) they have no access to high quality educational services and a motivating school environment. Good health, a skill-enhancing home environment and a good school are important conditions of skill acquisition. The more sustained the deficits are in this respect, the weaker skills we can expect.

1. Health. Pain, fatigue and stress associated with bad 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, letting them spend less time in the company of their peers and giving them fewer opportunities for sports and other activities that can help to develop their skills (*Currie* [2005], *Case–Lubotsky–Paxson* [2002], *Almond–Currie* [2011a]). Children with Attention Deficit Hyperactivity Disorder cannot concentrate as well, have less self-control[10] and are more impulsive, which is detrimental to their school performance (*Currie–Stabile* [2009]). 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–Paxson* [2005], *Palloni et al* [2009]).

Extensive international research shows that the children of less educated, poor families have a greater than average risk of contracting chronic diseases and of suffering accidents and injuries. Moreover, as parents are less likely to recognize the symptoms of disease; have less access to the right healthcare institutions due to insufficient information and transportation options; and are less able to afford the cost of care, their children have a more difficult time recovering. The health of children from low-income families is systematically worse than that of their higher-income counterparts, and this difference increases with age (*Case–Lubotsky–Paxson* [2002], *Currie–Stabile* [2003], *Currie* [2009]).

2. Home environment/parenting. Hundreds of activities, tools, aspects of the material environment and behavioral patterns combine to form the home learning environment. We single out two components: *1.* the availability of activities, objects, tools and environmental factors that directly or indirectly promote the child's cognitive development, and *2.* family parenting practices that guarantee the child's emotional stability (*Linver–Brooks-Gunn–Kohen* [2002]).

Just as the absence of environmental factors and activities that stimulate cognitive and linguistic development limits the development of the basic skills needed for learning, stressful family life and a lack of emotional support also have a negative impact on skill development and are detrimental to

10 *Moffitt et al* [2011], a groundbreaking longitudinal study that ran for over 30 years, provides strong proof that childhood self-control has a significant effect on life-changing events in adulthood.

school performance.

A family's human and material resources affect children's motivation and cognitive-emotional development via a series of complex family mechanisms. A considerable part of the disadvantages that are due to poverty are transmitted to the following generation through these channels. The literature on the inner dynamics of family life offers two theories to explain the hidden mechanisms of intergenerational poverty transmission. The *human capital theory*[11] claims that a low level of parental investment is responsible for the negative impact of the parents' poverty and lack of education on children's development. The effects of poverty and lack of education 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*[12] asserts that economic hardship or the loss of a job influences the child's development by means of the parents' mental state. As the parents' mental state affects the parent-child relationship and the parenting methods used in the family, it has a major impact on the child's development. The two classes of explanations are, to some extent, competing theories, but they complement each other in many respects.

3. *School quality.* Two main factors can make a school a "high quality" institution: effective teachers who get results and mutually motivating classmates who can learn from each other. In fact, there is a close correspondence between teacher quality and the composition of the peer group. Although measuring teaching quality can be a complex challenge, a number of innovative studies in the past two decades have shown convincingly that teacher performance plays a definitive role in the students' school performance. These studies assess teaching quality using a variety of methods: some measure observable features, such as the results of teacher skill tests (*Ferguson* [1998]), others measure student performance using value added models (*Rivkin–Hanushek–Kain* [2005], *Chetty–Friedman–Rockoff* [2011]), yet others compare the outcomes of current and less up-to-date pedagogical methods in the classroom (*Wenglinsky* [2001], *Schacter–Thum* [2004]). Their results are clear: high quality teaching is one of the main catalysts of good student performance.

The composition of the peer group also has a significant effect on students' performance. If any kind of social mechanism causes children with learning problems to cluster in one part of the 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–Kranton* [2002], *Bishop et al* [2003], *Fryer–Torelli* [2010]). A number of studies argue that high performance peer groups enhance, while low

11 *Leibowitz* [1974], *Becker* [1981a], [1981b], *Becker–Tomes* [1986], *Haveman–Wolfe* [1995], *Mayer* [1997], *Mulligan* [1997], *Kalil–DeLeire* [2004], *Guryan–Hurst–Kearney* [2008], *Gould–Simhon* [2011], *Kaushal–Magnuson–Waldfogel* [2011], *Phillips* [2011].

12 *Elder* [1974], *Lempers–Clark-Lempers–Simons* [1989], *McLoyd* [1990], *Conger et al* [1992], [1993].

performance peer groups inhibit individual learning performance (*Ammermueller–Pischke* [2009], *Hanushek et al* [2003], *Hanushek–Kain–Rivkin* [2009]).

The composition of the teaching staff typically mirrors peer group composition. School systems that isolate children who underperform or have learning difficulties – this group generally has a larger percentage of children from poor and less educated families – in special schools or classrooms also tend to have better teachers working in the schools attended by middle-class children, who are easier to teach, while less competent teachers remain in the schools attended by poorer children.[13] This indirect mechanism clearly diminishes the performance potential of disadvantaged students.[14] A number of cutting-edge studies have demonstrated the consequences of counterproductive teacher distribution in segregated school systems (*Hanushek–Kain–Rivkin* [2004], *Clotfelter–Ladd–Vigdor* [2005], *Jackson* [2009]).

The HLCS data allows for measuring these three mechanisms to varying detail.

1. We used two items of information that have proved useful in previous research: birth weight and health self-evaluation, to calculate a first estimate of student *health.*

Birth weight data has been collected regularly in a number of countries for over a century (*Ward* [1992]), providing information on the relationship of birth weight and later outcomes. Birth weight is one of the most important indicators that characterize the circumstances of pregnancy and fetal development. Children born with a low birth weight – operationalized as 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 a year of school; and have lower test scores (*Breslau et al* [1994], *Hack– Klein–Taylor* [1995], *Reichman* [2005]). Besides correlations, some studies have demonstrated the causal effects of low birth weight on level of education, employment chances and incomes (*Currie–Hyson* [1999], *Behrman–Rosenzweig* [2004], *Black–Devereux–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

13 This selection process is motivated by a mechanism of wage equalization. If there is no wage difference between the two types of classroom situations, then teachers who have better alternative employment options – who are also the better educators – will choose a workplace with a better workload/wages balance (*Kertesi–Kézdi* [2005]).

14 Naturally, this does not mean that a well-designed experiment (with additional resources) cannot achieve good results with disadvantaged students in segregated schools. The assessments of the Harlem Children's Zone (*Dobbie-Fryer* [2011a]) and the Knowledge is Power program (*Angrist et al* [2010]) report positive results. However, this does not imply that, in general, the phenomenon of counterproductive teacher selection is not typical of segregation, or that this phenomenon does not decrease the performance of disadvantaged children.

country is, the greater the statistical probability of low birth weight.[15] The relationship between the poverty of the parents and the low birth weight of the child has various causes. The mothers' inadequate nutrition (inadequate protein, vitamin and mineral intake), prior infections, greater prevalence of environmental hazards in or near the home, poverty-related stress,[16] limited access to healthcare institutions – all of these factors play a role (*Rosenzweig–Schultz* [1982], *Hack–Klein–Taylor* [1995], *Cramer* [1995], *Schonkoff–Phillips* [2004] Chapter 8, *Paul* [2010], *Currie* [2011]).

The second variable that we used to characterize health was self-rated health[17] of the surveyed students. The respondents evaluated their own health on a scale of one to four (excellent/good/adequate/poor). This variable, which is widely used in the literature, is strongly correlated both with medically diagnosed chronic conditions (*Case–Lubotsky–Paxson* [2002]) and with the indicators of the parents' social status (income and education). Poorer children generally tend to have worse health, and this is reflected in their self-evaluations, or, in the case of younger children, in their parents' subjective evaluations (*Case–Lubotsky–Paxson* [2002], *Currie–Stabile* [2003], *Case–Fertig–Paxson* [2005], *Currie* [2009] Table 1).

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

Early childhood experiences and family interactions relating to books and other written texts play an exceptionally important role in children's cognitive development. As Zita Réger points out: "Psychological studies of reading in educated families show that children gain a great deal of experience in the use of written and printed texts even before reaching school age as a result of conscious teaching and practicing with parents and older children. In the course of this process, they acquire a number of skills related to language use and cooperative interaction which are fundamental to the later acquisition of reading and writing skills, as well as to communication in school based on written and printed texts." (*Réger* [1995], 103). Regular bedtime storytelling sessions and parent-child interactions centered on browsing children's books together (including

15 Figure 1 in *Behrman–Rosenzweig* [2004], which provides birth data from 112 countries in the 1990s, has a very high dispersion (5-30 %) for the incidence of low birth weight. In developed countries, 6-9 % of all infants are born with a birth weight under 2500g, while in Third World countries this rate often reaches 15-20 %. The birth weight of the students in the HLCS sample is directly comparable to these data, as the bulk of the students were born at the same time (1990-1992). The incidence of low birth weight is 8 % in the full sample. However, the average conceals dramatic socio-economic and educational differences: while the incidence of low birth weight is 5-6 % for mothers who have completed secondary school, it is double that (14.4 %) for mothers who have completed no more than 8 years of schooling.

16 Increased frequency of smoking during pregnancy also plays an important role (*Meara* [2001]). Two factors implicated in this phenomenon are economic hardship, which can cause stress, and a lack of education, which can prevent awareness of the risks.

17 In this study, we used the question asked in the 2006 fall wave. The participating students were 14-16 years old at that time, and the modal age was 15.

picture books) are by far the most important ways in which toddlers and kindergarten-age children acquire such experiences. The amount of reading- and writing-related experiences in early childhood has a crucial effect on the child's basic skills prior to school enrollment (*Heath* [1983], *Réger* [1990], *Neuman* [1996], *Sénéchal et al* [2001], *Dickinson–Tabors* [2001], *Raikes et al* [2006]). There are two independent sources of data on the frequency of bedtime storytelling sessions during kindergarten age in the HLCS, as the question was posed separately to both the parents and the children in different parts of the interview. The only other question on early childhood skill development activities in the HLCS[18] that had a significant correlation with 8th grade test scores asked whether children hiked/did sports with their parents. Consequently, this was the only information that we included in the following.

We measured students' current home environment and the families' parenting with the HOME scale, which is widely used in the developmental psychology literature. The HOME *(Home Observation for Measurement of the Environment)* index is an instrument used to assess the developmentally relevant features of a child's home environment. Different versions have been designed for various age groups: infants, kindergarten age children, elementary school children and adolescents. The HLCS was the first attempt to bring the adolescent (ages 10-15) version of the HOME index (*Bradley et al* [2000], *Mott* [2004]) to bear on a large sample in Hungary. "In designing the scale, the researchers' assumption was that the aim of development in early adolescence is for the individual to become a healthy and useful member of society, capable of meeting societal expectations regarding work, personal relationships and responsibility. To accomplish this, the adolescent must show development in five different areas. He or she must become capable of establishing emotional relationships; of developing a coherent and positive self-image; of making informed decisions; he or she must acquire the skills necessary to participate in the labor force; and learn to develop personal values and convictions. Research in developmental psychology claims that development in these areas is facilitated by tolerant and attentive parental behavior, the encouragement of an example set by adults, devices that promote both recreation and learning, a safe and aesthetically pleasing environment, cultural experiences involving the entire family, activities that recur regularly and a family environment that encourages independence while maintaining some form of supervision. The HOME scale attempts to assess these features." (Medgyesi [2007]).[19]

Recent research has shown that the home environment and parenting as measured by the HOME

18 These questions concerned activities such as playing board games, cooking, drawing, going to the theater and the cinema, going hiking and doing sports together with the parents.

19 The details of the survey can be found on the NLSY website: NLSY79 Child and Young Adult User's Guide 2002 (http://www.bls.gov/nls/y79cyaguide/nlsy79cusg.htm). Appendix A. NLSY79 Child HOME-SF, 184–185 and 192–194. The HLCS followed this procedure in every detail *(Medgyesi* [2007]).

scale are strongly related to children's school readiness and later school performance *(Crane* [1996], *Guo–Harris* [2000], *Linver– Brooks-Gunn–Kohen* [2002], *Brooks-Gunn–Markman* [2005], *Todd–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) used in the National Longitudinal Study of Youth (NLSY). The short version is composed of 27 items, and assesses two subscales: cognitive stimulation and emotional support.[20] We used the two subscale variables created from the appropriate items. As a supplemental measure of the family's human resources, we also included a key variable of the PISA studies (the number of books in the home) and information on the availability or lack of an Internet connection.

We describe the variables that characterize students' health and home environment and parenting in *Table 9.* Table A6 of the appendix shows the basic data on Roma and non-Roma students for these variables.

Table 9
The indicators of the transmission mechanisms

Name of variable	Definition of variable
HEALTH	
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 *self-evaluation,* is adequate or poor according to a four-part scale (poor / adequate / good / excellent): yes/no (modal age: 15)
HOME ENVIRONMENT	
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 did sports together with the parents while the child was in kindergarten: yes/no (child's response)
Cognitive HOME index	The subscale of the HOME index (a synthetic variable characterizing the home environment) for 15-year-olds that measures cognitive stimulation
Emotional HOME index	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

3. In contrast to health home environment we relied on an indirect method to measure the relationship between *school quality* and the test score gap. We calculated the gap between students who studied in the same school and same class by estimating regressions including school/class

20 We present the items that form the basis of the subscales in the *Appendix.*

fixed effects. Recall that the HLCS is based the administrative NABC database, which contains the students' school and class identification numbers besides their test scores. The HLCS sample was large enough (and followed a multi-stage sampling) so there are a sufficient number of classmates in the sample.

When interpreting the results, we can think of the regression estimates of the "Roma" coefficient in the equations without school fixed effects measuring the differences between randomly selected Roma and non-Roma students in the sample. The "Roma" coefficient in the equations that do include school fixed effect measures the gap between Roma and non-Roma *classmates*. The difference of the two estimates measures the test score gap between Roma and non-Roma students who are *not classmates*. This component of the ethnic performance gap presumably incorporates the consequences of school segregation on the Roma students, which has a performance-reducing effect. If a large number of Roma students, whose average performance is lower, are taught in separate schools or classrooms (because of residential segregation, student mobility between school districts, or local government school policies that exacerbate segregation), then – for the reasons given in the previous chapter – the education that these children receive is typically and systematically21 of inferior quality: they can expect to have low-performing teachers and a peer group that is less likely to be conducive to learning.

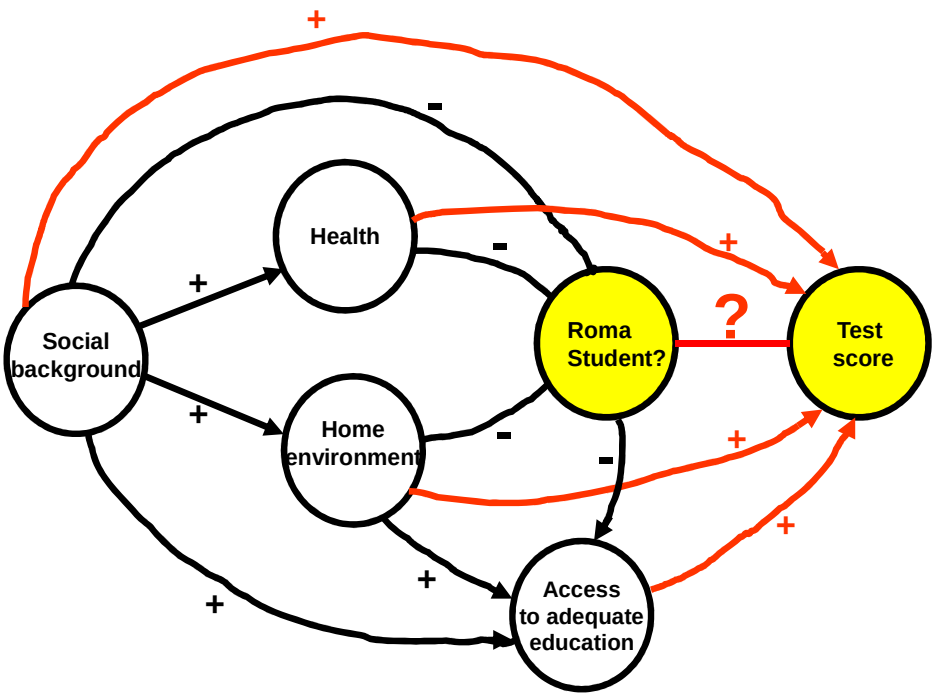
To be sure, we cannot rule out the possibility of selection bias. If local school systems contain elite, mixed and segregated schools, then it is possible that more of the lower-ability Roma children will attend a segregated school, while more of the higher-ability Roma children attend a mixed school and, conversely, the higher-ability non-Roma children have a higher probability of attending an elite school with no Roma students at all, while the lower-ability non-Roma children are more likely to attend a mixed school. Such selection patterns can increase the ethnic test score gap between schools even *without* a decrease in school quality. As we cannot eliminate the potential effects of selection bias, we give an upper estimate of the performance-reducing consequences of segregation by including school fixed effects. However, we have partially controlled for the heterogeneity of family background with a rich set of control variables, and we thus hope that the selection bias is not overly large.

We enter the intervening variables introduced in the previous chapter into the equations used to predict the test scores. We continue to estimate reduced-form equations, but we take the causal

21 Although some boards of education occasionally succeed in improving the instruction of children stuck in segregated schools with targeted programs and special resources, segregated schools typically and systematically have to face problems caused by counterproductive teacher distribution and the performance-decreasing effect of a low-performing peer group.

directions suggested by the theoretical literature into account, as illustrated by the schematic model in *Figure 5*. For the sake of simplicity, let us think of the theoretical features as one-dimensional variables *pointing in a single direction.* The family's socio-economic circumstances can be good or bad;[22] the child can have good or bad health; the learning environment can be more or less conducive to the development of the child's abilities;[23] the chances of access to high quality education can be good or bad; the test scores can be high or low. The simple lines (not arrows) connecting the variables designate a simple correlation, arrows designate a cause-and-effect relationship, and the plus and minus symbols indicate the sign of the relationship. The Roma students' school deficits are embedded in this causal network. We have already seen – and the expanded model will make it even clearer – that a naive interpretation of the ethnicity–test score relationship as a simple two-variable relationship breaks down if we attempt to account for it in the context of the generally valid causal relationships that determine school performance.

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



The family background variables that characterize a family's socio-economic circumstances are predetermined variables in the sense that they precede the other variables. These are typically the factors (income and wealth, parental education and employment, place of residence) that influence children's health, the home learning environment (the material conditions and activities that

22 Income per capita and parental education can be high or low; families can live in areas that offer few or many educational-cultural opportunities.
23 HOME cognitive subscale values can be high or low.

facilitate children's development), and the chances of access to quality education; and not vice versa.[24]

The chances of access to quality education are regulated by selection mechanisms based on social and ethnic cues and basic skill levels.[25] A number of factors are at play here: significant differences in social composition and income between towns; residential segregation within towns and villages; large-scale student mobility between school districts;[26] selective admissions to better schools; and occasionally, local school policies that exacerbate segregation. The cumulative effect of these factors is that disadvantaged and poor children with initially weaker basic skills (this includes Roma children) are highly likely to study at different schools and in different classrooms than children from families living in better socio-economic circumstances who start their school careers with stronger basic skills (and are predominantly non-Roma) (*Kertesi–Kézdi* [2005], [2009] and [2012a]). As we argued in the chapter on the transmission mechanisms, negative peer group effects and counterproductive teacher placement in a segregated school environment reduce the chances of access to quality education for disadvantaged students.

Although *Figure 5* presents the causes of school deficits in a form that resembles a structural model, we continue to estimate reduced-form equations in the following. Variable entry order follows the order of the schematic model. We start out from a two-variable relationship. Next, we enter the variables representing health, the home learning environment, and school/class fixed effects into the equation. Finally, we enter the family background variables that characterize the family's socio-economic circumstances. The latter can affect the test scores both *directly* and *indirectly*, through the student's health, learning environment, and access to high quality education. Our aim is to determine to what extent the inclusion of a given factor or group of factors reduces the test score gap measured by the "Roma" parameter. The interconnected nature of the intervening variables prevents us from isolating the effects of health, the learning environment and the school, but we can give a reliable estimate of their *combined effect.*

As *Table 10* demonstrates, the bulk of the raw test score gap *disappears* (over 90 % of the reading

24 There are also plausible examples of reverse cause and effect relationships. For instance, if a seriously ill child requires intensive home care, this can be why one of the parents does not work outside the home, and *that is why* the family has a low income, etc. However, we do not think that these reverse cause and effect relationships are representative.

25 The children of poor and less educated parents start elementary school with more difficulties (weaker basic skills) than their middle-class counterparts: *the skills and dispositions required for school* are at lower than average levels (*Heath* [1986], *Réger* [1990] and [1995], *Lee–Burkam* [2002], *Józsa* [2004], *Brooks-Gunn–Markman* [2005], *Neuman* [2006]).

26 There is a high level of student mobility between school districts in Hungary due to a system of free school choice. This process is highly selective. 8th grade student data from the 2006 NABC reveal that the ratio of students attending school outside their own school district is only 10-15 % among the children of less educated mothers (who completed grades 0-7 or 8), 33 % among the children of mothers with a high school diploma, and 50 % among the children of mothers with a university diploma.

and over 80 % of the mathematics test score gap) if we account for the variables corresponding to the social mechanisms. That is, if there were no difference between Roma and non-Roma students in terms of health, if they had the same degree of access to the resources, tools and activities that stimulate skill development in their home environment, and if they had the same chances of access to high quality education, then their 8th grade school performance would exhibit no difference, or only a minor difference. There would be no gap at all in reading comprehension skills, and the difference in mathematical skills would be small. The school deficits of Roma students are due exclusively to *well-defined social mechanisms,* and *not* to *ethnicity*.

Table 10
The magnitude of the residual ethnic test score gap after accounting for the transmission mechanisms

	Reading comprehension			Mathematics		
Roma	-0.97	-0.07	-0.05	-1.05	-0.18	-0.15
	(0.05)**	-0.07	-0.07	(0.05)**	(0.07)**	(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^2	0.06	0.67	0.68	0.07	0.68	0.69

Standard errors in parentheses.
* Significant at 5 %, ** Significant at 1 %.
Note: see detailed results in Table A7 of the appendix.

It is difficult to establish the relative magnitude of these mechanisms, because – as mentioned above – the relevant variables are highly correlated. If we enter them into the equations as separate factors, then their effect will appear *greater* than it is, as they incorporate the effects of the other variables with which they are correlated. And if we enter them into the equations last, after entering the other variables, then their effect will appear *smaller* than it is, as their effect will already partially have been accounted for, by the variables entered earlier with which they are correlated. The truth must lie somewhere between these estimated minimum and maximum values. We measured the effects of every transmission mechanism using both methods; the results are shown in *Table 11.* As before, we measured the effects by examining how much the inclusion of a new factor in the equation decreases the residual ethnic test score gap.

Table 11
The relative strength of the transmission mechanisms' effects

	Reading comprehension		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

Although the range of the estimates is rather broad, it seems clear that the home learning environment and the chances of access to high quality education play a decisive role in the test score gap between Roma and non-Roma students. Concerning the family background variables that characterize a family's socio-economic circumstances, a comparison of *Tables 8* and *9* reveals that they are related to test scores almost exclusively through these transmission mechanisms; their effect independently of these channels is small.[27] The results suggest that the test scores of Roma students are worse *because* they have little access at home to resources and activities that promote skill development. Similarly, the results suggest that the test scores of Roma students are worse *because* they have no access to good schools due to residential disadvantages and the selective mechanisms of the school system.

The bad socio-economic circumstances of Roma children are also reflected in their relatively worse health. However, health plays a less important role in teenage test results than the home environment and the quality of instruction. Although childhood health problems are important in terms of later life outcomes, their effect may be less apparent on the short term in school performance. Instead, they are likely to determine long-term outcomes by undermining adult health, reducing life expectancy and employment chances. Although we cannot corroborate this claim with the available data, a number of recent longitudinal and retrospective studies have reached a similar conclusion (*Elo–Preston* [1992], *Case–Lubotsky–Paxson* [2002], *Case–Fertig–Paxson* [2005], *Smith* [1999], [2009], *Marmot–Wilkinson* [2006], *Strauss–Thomas* [2008] Chapter 4). Thus, it is likely that poor childhood health has similar effects in Hungary.

Having observed that differences in the home environment and schools are very strongly related to the test score gap, our next question is: to what degree do the Roma students' socio-economic disadvantages explain their deficits in these two factors? In the case of the home environment indicators, we attempt to draw conclusions about the way family dynamics. How are family dynamics influenced by adverse socio-economic circumstances (residential disadvantages, income poverty and low parental education), and what, if any, role do independent ethnic factors play? In the case of school disadvantages, we attempt to understand how the school system's selection mechanisms work. What role do Roma students' social disadvantages and skills deficits (due to the home learning environment) play in the observed large-scale school segregation? And once we have

27 The full effect estimated using the equations in *Table 4* is 0.74 (reading) and 0.73 (mathematics) units of standard deviation (i.e. the difference between the raw gap and the OLS estimate). After accounting for the transmission mechanisms – see the differences between columns 2 and 3, and between columns 5 and 6 in *Table 6* – the residual effect is only 0.02 (reading) and 0.03 (mathematics) units of standard deviation.

accounted for these factors, what is the magnitude of the ethnic residual (an index for ethnic selection in the school system)?

Home environment

Table 12 shows two measures of the ethnic gap in the home environment indicators (see *Table 9*): the raw gaps and the gaps corrected for the family background variables (see *Table 7*).

Table 12
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^2
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
Cognitive HOME index	-1.118	(0.051)**	–	9056	0.09
	-0.080	(0.070)	yes	9056	0.70
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 %.
Note: see detailed results in Tables A8.1-A8.3 of the appendix.

The raw difference is significant for the majority of the variables: compared to non-Roma parents, nearly 30 percentage points fewer Roma parents regularly told their kindergarten-age children bedtime stories;[28] the cognitive HOME index for 15-year-olds is more than one unit of standard deviation lower for Roma families than non-Roma families; over 60 % of Roma families have no or only a few books at home, as compared to less than 10 % of non-Roma families. However, these differences are overwhelmingly due to long-term socio-economic disadvantage: nearly 90 % of the

28 Although the parents generally recall a lower frequency of bedtime stories than the children, the Roma/non-Roma *difference* is approximately the same. The non-Roma/Roma ratio of frequent storytelling is 64.6/35.0 % according to the parents (difference = 29.4 %) and 48.4/21.3 % according to the children (difference = 27.1 %).

difference in the frequency of bedtime storytelling, over 90 % of the difference in the cognitive HOME index measurements, and over 50 % of the difference in the number of books in the home would disappear, if the non-Roma majority students were to live in *similarly bad* circumstances as the Roma students.[29]

Similarly to our previous analysis of the test score gap along the distribution of family background, we can examine the ethnic differences in the home environment measures along the distribution of family background, too. We use the previously introduced synthetic variable – the family background index – for that purpose. Analogously to the previous analysis, we divide the range of the family background index (the linear combination of family income, poverty, parental education and parental employment) index into 10 equal intervals, and estimate the mean values of the home environment variables for the Roma and non-Roma students. Again similarly to the previous analysis, we restrict the estimates for the Roma students to the 0–0.6 range, because the subsample is too small to be meaningful over values of 0.6. The gray zone marks the 95 % confidence intervals (within ± 2 standard errors of the mean). This simple method of measurement is preferable to the OLS estimates presented in *Table 8* for three reasons: it contains no linearity constraints; it allows us to demonstrate potentially heterogeneous effects; and – most importantly – it measures the relationship along the full distribution of the variables, not only around the averages.

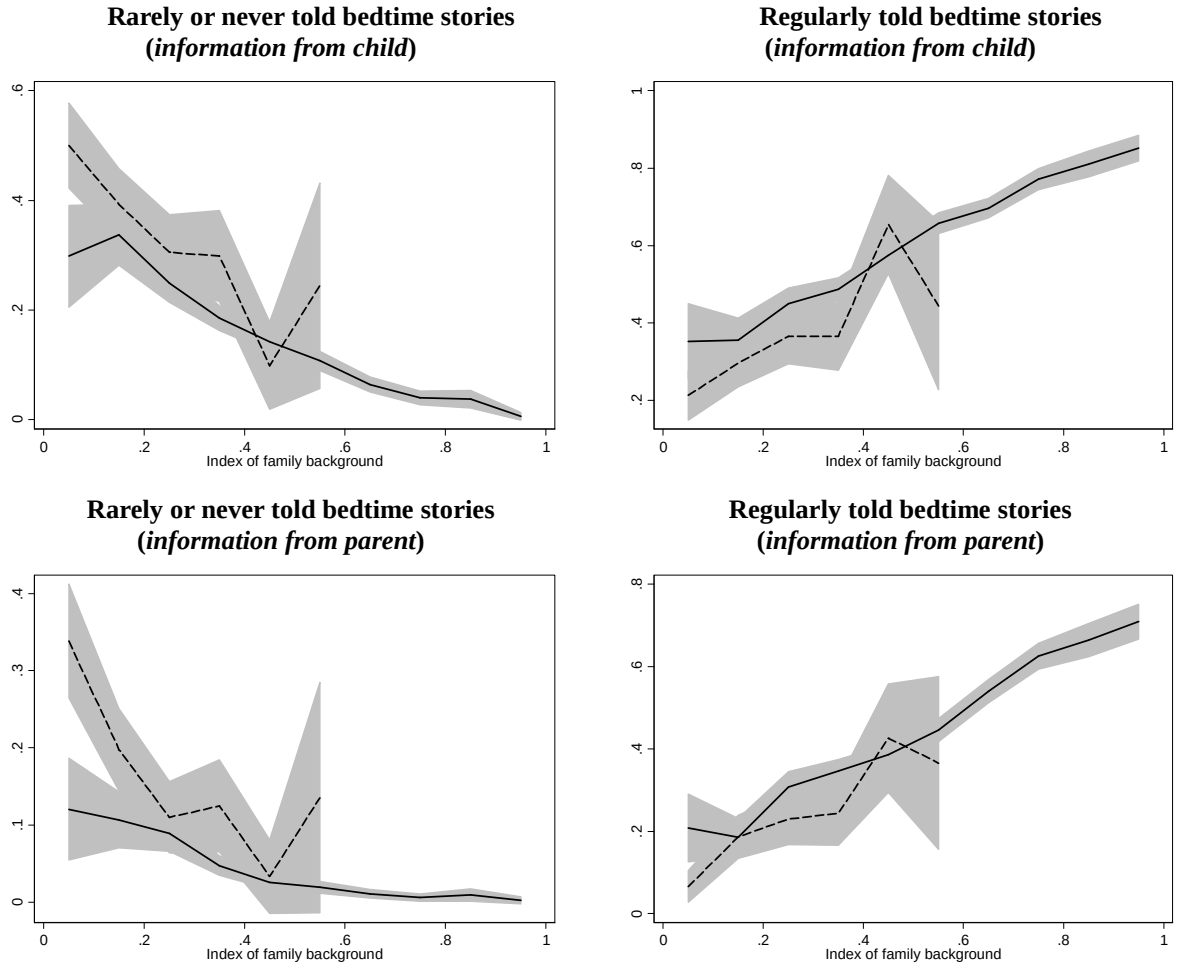
We present our results in the following two figures. Social and ethnic differences in the frequency of bedtime storytelling to *kindergarten age* children are shown in *Figure 6*, and the differences in the four indicators of the *teenage* home environment (cognitive and emotional HOME index, number of books and Internet access at home) are shown in *Figure 7*. The retrospective data on the frequency of bedtime storytelling was drawn from separate reports by the parents and the children.

29 This is how we interpret the corrected differences because non-Roma students make up a much larger part of the sample than Roma students (the ratio is approximately 9:1), and we estimate the parameters mainly using the non-Roma student data.

Figure 6

The probability of bedtime storytelling as a function of the family background index

Solid lines: Non-Roma. Dashed lines: Roma. The gray zone marks the 95% confidence intervals.



Figures 6 and *7* reveal, *first*, that the indicators with a significant association with test scores are strongly related to the family background index. There are vast differences in these factors[30] between rich and poor, educated and less educated, privileged and disadvantaged families. Only 20-30 % of the most disadvantaged students were told bedtime stories regularly in early childhood, compared with 70-80 % of children from the highest social status 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 units of standard deviation. 70 % of the poorest and least educated families have either no or very few books, which is true of none of the highest social status families. Less than 5 % of the poorest families had home internet access in

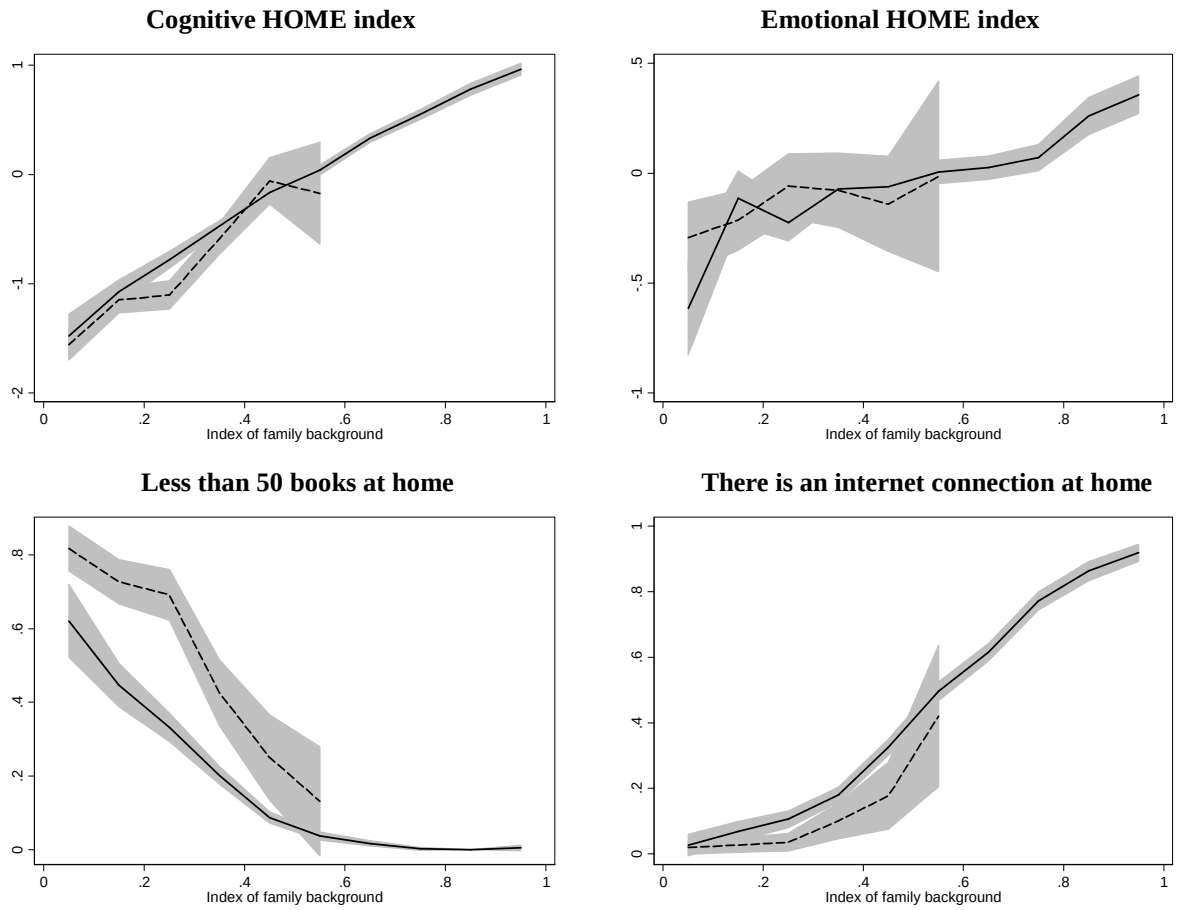
[30] All PISA studies agree that the number of books in the family's possession is an important determinant of student achievement (e.g. *OECD* [2010] 160). In the 2009 PISA study, 14 countries including Hungary participated in a supplementary survey in which parents gave an account of their parenting; one question asked whether parents read books with their children, and if so, how frequently. Children whose parents read to them daily or weekly had 25 % higher scores on the reading comprehension test at age 15 than children whose parents read to them less frequently or not at all (14 % after correcting for socio-economic background). The gap between the raw and corrected differences clearly implies that social differences have a strong influence on families' parenting methods (*OECD* [2010] 95).

2006, compared to 90 % of the wealthiest families.

Figure 7

The indicators of the adolescent home environment as a function of the family background index

Solid lines: Non-Roma. Dashed lines: Roma. The gray zone marks the 95% confidence intervals.



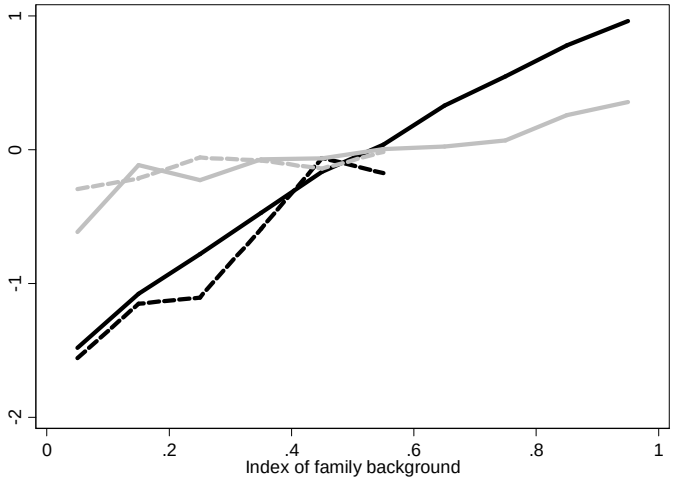
Second there are minimal, in many cases insignificant ethnic differences in the home environment indicators between families with comparable family background. Statistically, there is no difference in the regular bedtime storytelling and the cognitive and emotional HOME index graphs between the Roma and non-Roma students in the range with an adequately large sample of both populations. We are thus justified in arguing that, if Roma students were to live in *similarly good* circumstances as non-Roma students, these home environment indicator values would be at comparable levels. Although the ethnic residuum with respect to number of books owned and home Internet access would persist even in the case of full integration, it is very small in size compared to the raw difference (see Table A9 of appendix). As *Figure 7* demonstrates, the Roma/non-Roma gap narrows significantly as the family background index increases. In plain English, the results suggest that *ethnicity plays no demonstrable role* in the significant cognitive disadvantages associated with the parenting of Roma families; these disadvantages are entirely or almost entirely explained by the

parents' lack of education, poverty, and residential disadvantages. The children of *non-Roma* families living in similarly *disadvantaged* circumstances suffer similar cognitive disadvantages, while the parenting of *Roma* families living in *average* circumstances is not associated with such disadvantages.

Third, while we found a vast social gap in the HOME cognitive subscale scores, the differences in the emotional subscale scores were relatively small. Despite significant inequalities in socio-economic circumstances, high- and low-income families generally *do not differ greatly in their capacity to provide emotional security to their children (Figure 8).*

Figure 8
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



This is a surprising result, considering that the bottom third of society faces serious economic hardship, and that unemployment and economic hardship represent a major source of stress for families living in bad 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. The difference is that parents living in poverty are less able to create a learning environment (objects and activities) that is conducive to their children's skill development. Furthermore, as in the case of cognitive stimulation, *there are no ethnic differences in emotional security*. Roma families provide their children with the same level of emotional support as non-Roma families living in similar circumstances.

Although emotional security does not exhibit a significant relationship with poverty, it is closely connected to *family structure.* As *Table 14* shows, two-parent families are able to provide the

highest, and single mothers the lowest levels of emotional support. The difference between these two family types accounts for 70 % of the standard deviation of the HOME emotional subscale scores.[31] We emphasize this because the literature on family structure (*McLanahan–Sandefur* [1994], *Amato* [2005], *McLanahan–Percheski* [2008], *Sweeney* [2011]) claims that lack of family cohesion plays a major role in the reproduction of poverty. Thus, the breakdown of black family cohesion in the last 50 years – primarily since the 1980s, coinciding with a jump in social and economic inequality – is an important factor in the social disadvantages of black children in the United States. According to 2009 data from the US Census Bureau, 54 % of black children under the age of 18 live in a single-parent household (the overwhelming majority are raised by a single mother), compared to only 21 % of white children.[32] By contrast, *Table 13* shows that there is no such difference in the structure of Roma and non-Roma families in Hungary: the percentage of Roma children growing up without a father is well under 20 %, compared to approximately 20 % of non-Roma children. In the past 20-25 years, the Roma population in Hungary has experienced a crisis at least as severe as that of the black population in the United States, but the unprecedented scale of social exclusion affecting the Roma parent generation (*Kertesi* [2000], [2005]) has not – at least, not yet – caused a mass breakdown of Roma family cohesion. This must serve as a warning signal to Hungarian social policy: the situation may deteriorate further, and, once destroyed, it is very difficult to rebuild family cohesion.

Table 13
Family structure and emotional security

Family structure	Roma children (%)	Non-Roma children (%)	Cognitive HOME-index average	Emotional HOME-index average
Both the mother and the father are biological parents	73.7	68.2	0.09	0.14
The mother is a biological parent, the father is not	5.9	8.2	-0.16	0.15
The mother is a biological parent, father is not present	16.1	20.4	-0.14	-0.55
Other family types	4.4	3.3	-0.40	0.14
Total	100.0	100.0	0.00[a]	0.00[a]
Roma / non-Roma gap	–	–	–1.02	–0.18

[a] *standard deviation* = 1.00.

We also need to discuss *why* children living in adverse circumstances have suboptimal access to

31 The comparable difference is much smaller in the case of the HOME cognitive subscale, barely exceeding 20 %.
32 http://www.census.gov/population/www/socdemo/hh-fam/cps2009.html US. Census Bureau: America's Families and Living Arrangements: 2009, Table C9.

objects, activities and experiences that promote their development in their home environment.

1. The most obvious cause is *income poverty*: poor families are less able to afford the objects, tools and services that promote skill development than wealthier families. A new study of consumer expenditure data from the United States found that families in the top income quintile spent more than six times as much on *enrichment expenditures*[33] than families in the bottom income quintile in the middle of the 2000s (*Duncan–Murnane* [2011b] p. 11, *Kaushal–Magnuson–Waldfogel* [2011]).[34]

2. Middle-class families and the families of children with poor and less educated parents differ in terms of *parental time usage*. As no Hungarian studies were available on this topic, we rely on international research. Time use surveys from a number of developed countries demonstrate that less educated parents spend significantly less time[35] with their children – even though they are less likely to be employed and earn less – than more educated parents (*Sayer–Gauthier–Furstenberg* [2004] p. 1164, *Guryan–Hurst–Kearney* [2008] p. 35, *Ramey–Ramey* [2010] p. 137).[36]

This phenomenon has diverse causes. Some activities require both money and material resources, and if they do not fit into the family budget (see point 1), the parents do not invest complementary resources (time) either. Another reason may be that less educated parents are not as proficient in the aspects of parenting essential to skill-building activities. As their own reading comprehension skills are generally weak, they find it difficult to read together with their children,[37] to tell bedtime stories, and to monitor their school-age children's academic progress. A third reason may be that less educated parents are unaware of the importance of such activities (regular bedtime storytelling, cooking or hiking together) for their children's school readiness and later skill development. Yet another reason may be that some poor families living in social isolation (residential and school segregation) belong to subcultures in which free time is used differently than in middle-class families. The children spend more of their free time together with other children of similar social status, which reduces the frequency of parent-child interactions (*Heath* [1984], *Lareau* [2003]). This

33 The cost of leisure activities, private lessons, transportation to school and after-school/leisure activities, childcare, school, textbooks. The differences are similar in magnitude if we examine specific objects, tools and services rather than expenditure items: books and newspapers, computers, sports equipment and activities, travel, electronics.

34 Moreover, this gap widens significantly when income inequality increases (*Duncan–Murnane* [2011b] p. 11, Figure 1.6).

35 Time spent with children includes primary childcare activities (breastfeeding, rocking-putting to sleep, feeding, changing diapers, visiting doctors, providing physical care, etc.), educational childcare activities (reading aloud, telling bedtime stories, assisting with homework, attending children's events, etc.), and recreational childcare activities (playing with children at home or outdoors, participating in the child's sports, music or dance activities, trips to the theater, cinema or the zoo, going on walks, etc.).

36 According to data from the United States, a working mother with a university diploma spends an average of 6.5 more hours per week with her child than a working mother without a high school diploma. These results have been corrected for differences in the number of children, marital status and the children's age (see *Guryan– Hearst–Kearney* [2008] p. 35).

37 "Parents [...] who reported a low level of literacy initially found themselves struggling with reading and not enjoying the experience of reading together with their child." (Neuman [1996] p. 510)

keeps children confined within their residential area – the younger they are, the more so – and limits their opportunities to acquire information about the outside world. *Phillips* [2011] relied on representative time use data from the US to demonstrate that the difference between the bottom and the top income quintile in time spent in non-routine places is 4.5 hours for toddlers under 2 years of age and 3.7 hours for kindergarten age children. Cumulatively from birth to age 6, low-income children spend 1300 fewer hours away from their accustomed environment than children from the top income quintile (p. 217).[38]

3. There is also a difference in the quantity and quality of *parent-child interactions* in middle-class families and families with less educated, low-income parents. Less educated parents speak significantly less with their children; their vocabulary is smaller; they use fewer nouns, adverbs, adjectives and employ the past tense less frequently to express themselves; they address more commands and fewer questions to their children; and their parenting incorporates fewer encouragements and more discouragements than that of more educated parents (*Réger* [1990], *Hoff-Ginsberg* [1991], *Hart–Risley* [1995], *Huttenlocher et al* [2002], *Hoff* [2003], [2006], *Phillips* [2011]). The difference in linguistic experience plays a fundamental role in disadvantaged children's lack of school readiness. (*Neuman* [2006]).

4. Disadvantaged children have little access to books and printed materials in general, not only in their immediate family environment, but also in their residential area. Children living in villages and segregated neighborhoods seldom see street signs, advertisements, or storefronts which, on repeated viewings, promote spontaneous acquaintance with letters and numbers (*Neuman–Celano* [2001]. These areas are generally underserved by kindergarten, school and local public libraries. (*Neuman* [1999], *Neuman et al* [2001], *Neuman–Celano* [2004]). There has been little effort on the part of decision-makers in education in the past 20 years to facilitate disadvantaged children's access to books and digital culture either through the education system, or by improving the public library network.[39]

Education opportunities

Another key factor in the test score gap between Roma and non-Roma students is Roma students'

38 Menyhért Lakatos Elementary School is in the 8th district of Budapest; its student body is largely made up of children from less educated, low-income families. A few years ago, the principal carried out a survey among the newly enrolled first graders, asking how many of them had ever seen the Danube. The results were crushing: over half of the 6-7-year-olds had never been to see the Danube, despite living in a central district of Budapest.

39 A program that provides textbooks to disadvantaged children free of charge is an exception in this regard. However, the program has one major flaw: the children have to return the textbooks at the end of the school year. This makes it impossible to build on the knowledge acquired in previous years in a number of subjects (mathematics, IT, foreign languages, the natural sciences). Moreover, if the textbooks did not have to be returned, the program could help poor children acquire their own little home library, just like middle-class children.

lack of access to good schools. This lack of access is due to residential disadvantages and the school system's selection mechanisms. The majority of Roma students are taught in a classroom context in which the sheer quantity of unresolved pedagogical problems makes it very difficult to teach well. In order to demonstrate this, we combine the HLCS sample with the full 2006 NABC database. For every student in the HLCS sample, we establish *what percentage of the other students* in his/her 2005-2006 8th grade class[40] performed at levels 0 or 1 (i.e. unacceptably poorly) on the reading comprehension test.[41] We consider a class *difficult to teach* or *highly segregated by ability* if more than half of the given students' classmates perform below level 2 on the reading comprehension test. We find enormous differences in class-level segregation thus defined between Roma and non-Roma students. In Hungary today, 58.1 % of Roma 8th graders are in classes in which over half of their classmates can be considered functionally illiterate, compared to 17.7 % of non-Roma students. The raw ethnic difference is 40 %: Roma students are 40 % more likely to end up in a class that is highly segregated by ability, a class in which it is almost impossible for teachers to provide quality instruction[42] due to the excessive workload and the adverse student body composition. As a result, the majority of good teachers avoid these classes,43 especially as there is no compensation for the higher workload.[44]

We use linear probability models in *Table 10* to estimate the composition effect on this difference.

---

40 In Hungarian schools, a "class" designates a group of students who attend most classes together, in the same classroom; not to be confused with all the students in a given grade who attend the same school.

41 The test designers divided the synthetic variables of basic skills,which had a mean of 500 and a standard deviation of 100, into 5 skill levels (0, 1, 2, 3, and 4). The threshold is level 2 in both cases. Students who are not at skill level 2 in reading comprehension or mathematical-logical skills have no useful knowledge regarding these basic skills. Students performing under level 2 in reading comprehension cannot identify one or more pieces of information based on multiple criteria and cannot connect obviously similar information. They cannot recognize basic connections within the text, develop and apply simple categories or draw lower level conclusions using one or more parts of the text. They cannot recognize the main idea of the text, interpret a specific part of the text or identify the author's intent. Their background knowledge is insufficient to evaluate an element of the text. Students performing under skill level 2 in reading comprehension are practically illiterate: they may be capable of reading a text, but they cannot identify its contents or use its meaning. These students can be considered functionally illiterate.

42 There are certainly cases in which it is possible to create a good school for children with learning difficulties *despite such obstacles.* However, this is the exception rather than the rule. It is possible to swim against the current, but the current still makes it more difficult to move forward. It is no coincidence that we are more likely to find good performances where there is no need to struggle against the current.

43 *Scafidi–Sjoquist–Stinebricker* [2007] report that growing segregation (which reversed previous gains) in Georgia public schools between 1991/1992 and 2000/2001 caused qualified teachers to increasingly leave segregated schools. *Jackson* [2009] reports similar findings from the Charlotte-Mecklenburg school district in North Carolina, which ended school desegregation and the busing of students in 2002, causing the previously integrated local school system to become highly segregated within the space of a few years: "Schools that experienced an increase in the black enrollment share saw a decrease in the proportion of experienced teachers, a decrease in the proportion of teachers with high scores on their licensure exams, and a decrease in teacher value added." (*Jackson* [2009] p. 248).

44 *Hanushek–Kain–Rivkin* [2004] calculate that very high salary increases (25-40 %) would be required to noticeably reduce the number of teachers leaving schools that predominantly serve low-performing students (p. 350). On the other hand, the widely known estimates of *Antos–Rosen* [1975] demonstrate that teachers are willing to accept a lower salary in return for the opportunity to teach high-performing students. The principle of wage equalization obtains in both directions. *Boyd et al* [2003] use a two-sided matching model to estimate the definitive factors of teachers' workplace selection and schools' teacher selection decisions. They estimate that a 30 % increase in the proportion of minority students (approximately one unit of standard deviation) leads to a utility loss for teachers that could be offset by a wage increase of approximately 1.3 units of standard deviation.

In our thought experiment, we estimate the difference between Roma and non-Roma students' chances of ending up in a highly segregated class if non-Roma students were to live in the same socio-economic circumstances (and in the same home environment) as Roma students. The greater the composition effect, the more persuasive the claim that *school selection* in Hungary is *based on place of residence, social background and ability*. The greater the residuum, the greater the *school segregation by ethnicity*. As in previous chapters, we use the family background and home environment variables to measure the composition effects.

Table 14
The probability of attending a class highly segregated by ability (number of observations: 9056)

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

[a] Over half of all classmates performed below level 2 on the reading comprehension test.
[b] Standard errors in parentheses.
* Significant at 1 %.
Note: see detailed results in Table A10 of the appendix.

The OLS estimates in *Table 14* show large composition effects, but the ethnic residuum is also significant.[45] The fact that a Roma student is 40 % more likely than a non-Roma student to be in a class that is highly segregated by ability is largely – in 65-70 % – due to the parents' lack of education, poverty, disadvantaged residential area, and lack of access in the family environment to the resources, services and activities that promote skill development. The one-third residuum demonstrates that school segregation by ethnicity plays a non-negligible role.

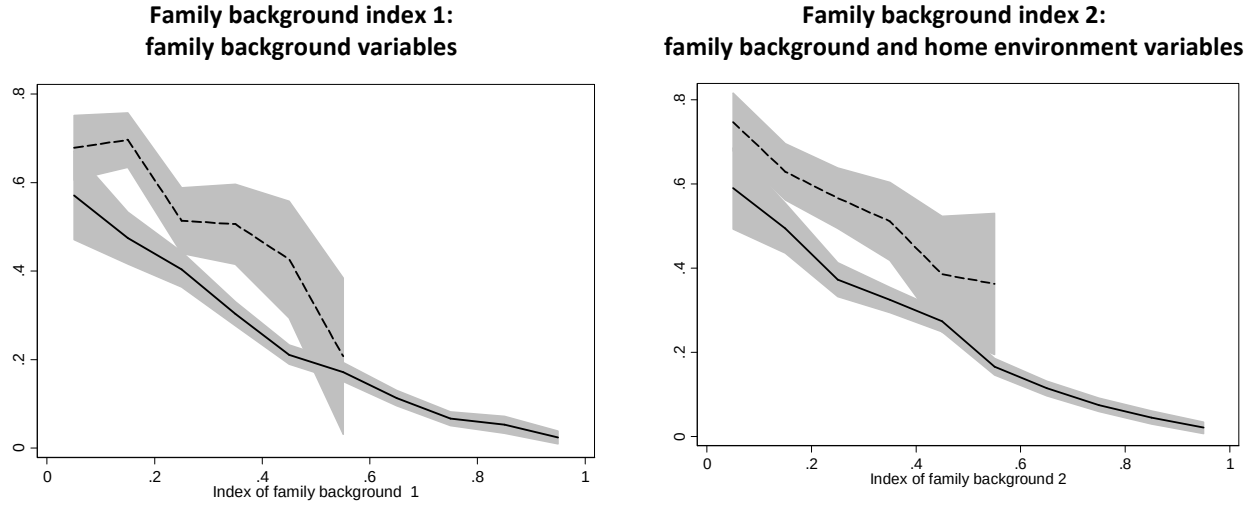
We use the previously introduced simple two-variable graphs to demonstrate that the statement above applies to non-Roma children living in socially disadvantaged circumstances as well as to Roma children, if to a somewhat lesser degree (*Figure 13*). We present the relationships without linearity constraints, along the full distribution of the synthetic measures of the social background variables. We have two synthetic variables on the horizontal axes of the two-variable graphs: the previously applied family background index on the left-hand side, and a new variable that includes the variables of the home environment as well as the family background variables[46] on the right. The variable on the vertical axis measures the likelihood that a given student attends a class in

45 See table A11 of the appendix.

46 To create the new synthetic variable, we used exactly the same procedure as in the creation of the family background index. We combined the variables in *Tables 3* and *5* into a one-dimensional synthetic variable: we estimated the linear combination of the regression parameters using the mean reading and mathematics scores, and then normalized the resulting values for a range of 0 to 1. Young people living in a worse home environment and socio-economic circumstances are thus in the range closer to 0, while young people living in better circumstances are in the range nearer to 1.

which more than half of his/her classmates performed at unacceptably poor levels (level 0 or 1) on the NABC reading comprehension test.

Figure 9
The probability of attending a class highly segregated by ability as a function of students' socio-economic background
Solid lines: Non-Roma. Dashed lines: Roma. The gray zone marks the 95% confidence intervals.



*Note: C*lass highly segregated by ability: over half of all classmates perform under level 2 on the reading comprehension test. Family background index 1: family background variables. Family background index 2: family background + home environment variables.

Figure 9 reflects a marked mechanism of selection by socio-economic background in the case of both Roma and non-Roma children. Students of low social status have a significantly higher chance of ending up in a segregated class than higher social status students, *independently of ethnicity*. Roma students face the additional disadvantage of ethnic segregation at all levels of the social status index. For the range of the social status index that the majority of Roma students fall into, we find that two-thirds of the large-scale school segregation of Roma children in Hungary is caused by poverty independently of ethnicity, and one-third by ethnic discrimination directed specifically at Roma students. Well-targeted anti-segregation policies should generally follow *color-blind* principles and sanction the segregation of poor and disadvantaged students in all its forms, independently of ethnicity. However, due to the non-negligible degree of ethnic selection, flagrant examples of school segregation do require *special attention* when segregation by skin color can be confirmed.

Summary and policy recommendations

This study quantifies the achievement gap between Roma and non-Roma students in East Central Europe and assesses the potential causes of the gap. Using the UNDP survey of 2011, the only

comparable data on the Roma spanning many countries, we show that the gap in the chances to get secondary education is substantial in all countries. When comparing young adults living with parents of comparable income and educational attainment, the gap drops by more than a half in most countries.

Using unique data from Hungary, we assess the gap in standardized test scores and show that it is comparable to the size of the Black-White test score gap in the U.S.A. in the 1980's; however, that gap has since narrowed significantly. To a large extent, these deficits explain Roma students' later lack of success on the labor market and the intergenerational transmission of Roma minority disadvantage. The skills gap emerges at a very early age, before enrollment in elementary school, and the differences as measured at the end of elementary school continue to increase in secondary school.

Social differences (in income, education and place of residence) account for a large part of the gap. If the non-Roma students lived in similarly bad socio-economic circumstances as the Roma students, or if the Roma students lived in similarly good circumstances as the non-Roma students, only a fraction of the gap would persist: one-fourth of the mathematics gap and one-fifth of the reading gap.

Roma children growing up in poor and less educated families and adverse residential conditions face a number of obstacles to learning: their health is worse than average, they do not have access to the resources that promote skill development in their home environment, and they have less access to good schools in the course of their education. These transmission mechanisms play a decisive role in turning Roma students' social disadvantages into school deficits. If there was no difference between Roma and non-Roma students in terms of health, if they had the same degree of access to the resources, tools and activities that promote skill development in their home environment, and if they had the same chances of access to high quality education, then their 8^{th} grade school performance would exhibit no difference, or only a minor difference. There would be no gap at all in reading comprehension skills, and the difference in mathematical skills would be small. The school deficits of Roma students are due exclusively to well-defined social mechanisms, and not to ethnicity.

Disadvantages in the home environment that play a key role in the school performance gap are largely explained by social differences. Ethnicity plays no additional role in the significant cognitive disadvantages associated with the parenting of Roma families; these disadvantages are fully or almost fully explained by the parents' lack of education, poverty, and residential disadvantages. The children of non-Roma families living in similarly disadvantaged circumstances suffer similar cognitive disadvantages, while the parenting of Roma families living in average

socio-economic circumstances is not associated with such disadvantages. Parents who live in poverty are far less able to provide an environment (objects, tools, activities) that promotes the skill development of their children than their middle-class counterparts.

Despite significant inequalities in socio-economic circumstances, high- and low-income families generally do not differ greatly in their capacity to provide emotional security to their children. This is a surprising result, considering that the bottom third of society faces serious economic hardship; unemployment and economic hardship represent a major source of stress to families living in bad socio-economic circumstances. Ethnicity does not play a role either in cognitive stimulation or in emotional security: Roma families provide their children with the same level of emotional support as non-Roma families living in similar circumstances. Despite long-term poverty and an unemployment crisis that has lasted over twenty years, Roma family cohesion is still comparable to that of non-Roma families living in much better circumstances. This must serve as a warning signal to Hungarian social policy: the situation may deteriorate further, and, once destroyed, it is very difficult to rebuild family cohesion.

Another key factor in the test score gap between Roma and non-Roma students, in addition to the disadvantages of the home environment, is Roma students' lack of access to good schools. This lack of access is due to residential disadvantages and the school system's selection mechanisms. The majority of Roma students are taught in a classroom context in which the sheer quantity of unresolved pedagogical problems makes it very difficult to teach well. Roma students are 40 % more likely to end up in a class that is highly segregated by ability and difficult to teach, a class in which it is almost impossible for teachers to provide high quality instruction due to the excessive workload and the adverse student body composition. As a result, the majority of good teachers avoid these classes, especially as there is no compensation for the higher workload.

Students of low social status have a significantly greater chance of ending up in a class segregated by ability, independently of ethnicity. However, Roma students also suffer the effects of ethnic segregation. The magnitude of the two effects have a two to one ratio: two-thirds of the marked segregation of Roma children in Hungarian schools is caused by poverty independently of ethnicity, and one-third by ethnic discrimination directed specifically at Roma students.

The Roma/non-Roma school achievement gap is primarily due to poverty and associated disadvantages at home and at school. Aside from the phenomenon of school segregation, none of the causes of the achievement gap require a social policy intervention directed at the Roma minority in particular. The academic deficits and social exclusion of disadvantaged children, both Roma and non-Roma, should be remedied by universal and color-blind policies.

Interventions should clearly aim to prevent extreme poverty in families with children; income and family support policies should pay particular attention to struggling social groups. A number of circumstances warrant family support policies that help the poor more effectively.

First, income inequality has been on the rise since the 1980s in developed countries (*OECD* [2011]), which has had an outsize impact on children (*McLanahan* [2004], *Ferge–Darvas* [2010], *Duncan–Murnane* [2011a]) and threatens to divide societies even further, presaging the long-term exclusion of the next generation of those living in poverty.

Second, income effects on children's behavioral indicators – cognitive and non-cognitive skills, school achievement and educational advancement – appear to be non-linear: extreme poverty causes disproportionately great damage (*Brooks-Gunn–Duncan* [1997], *Duncan–Brooks-Gunn* [1997], *Barajas–Philipsen–Brooks-Gunn* [2007], *Kertesi–Kézdi* [2007]), while income increases that affect the entire income distribution scale produce greater improvement at the bottom of the scale than in the middle and at the top (*Dahl–Lochner* [2011], *Loken–Mogstad–Wiswall* [2012], *Black et al* [2012]). In light of the transmission mechanisms that translate socio-economic circumstances into school achievement, conditional cash transfer programs that comprise positive incentives to promote children's health and skill development seem particularly promising (*Fiszbein–Schady* [2009]).

Societies cannot always rise to the challenge of significantly alleviating poverty or preventing the emergence of mass poverty among families with children. However, targeted policy interventions can successfully reduce the skills gap of children who grow up in poverty and marginalized social groups. A minimal precondition to achieving this goal is to understand the complex transmission mechanisms, both within families and in the wider community, that lead to the reproduction of poverty – this is one of this article's most important claims. With adequate planning, it is possible to intervene in these mechanisms even if a weak economy or lack of political support prevents the long-term, large-scale alleviation of childhood poverty. The remaining part of our summary chapter will present the principles that the design of such programs and measures should follow as suggested by international research.

The perhaps most promising method of preventing failures at school is to provide children with an environment (objects, tools, activities, services) that facilitates their cognitive and linguistic development, and to promote complementary parenting methods (*Herczog* [2008], *Almond–Currie* [2011a]), *Heckman* [2011]). The most important principles of designing parenting interventions are the following.

1. As the cognitive skills deficits that can lead to a lack of success at school appear at a very early

age, interventions aiming to counterbalance them should target the youngest possible age group (*Rouse–Brooks-Gunn–McLanahan* [2005]).

2. These interventions must devote particular attention to the development of disadvantaged children's linguistic environment, cognitive and non-cognitive skills and motivation.

3. As research suggests that programs that aim to alter families' parenting directly often remain unsuccessful, these goals may be better achieved by universal, institution-oriented programs (nursery schools, kindergartens, Sure Start centers) (*Brooks-Gunn–Berlin–Fuligni* [2000], *Waldfogel* [2006], *Furstenberg* [2011], *Phillips* [2011]).

4. One method of effectively influencing parenting is for institutions to actively involve parents in their work, and thus help them recognize the significance of playful learning, literacy, books and regular storytelling to children's development. However, such changes in family behavior can only be achieved if parents are treated as partners in education. Activities that bridge the distance between the families and the institutions (*Havas* [2004]) and help build parenting skills should be included in the complementary elements of such a program. Furthermore, programs must pay great attention to strengthening family integrity, parental autonomy and parents' feelings of competence.

5. As such universal, large-scale programs reach the entirety of an affected age group, it is worth adding health education elements to such parenting-type interventions. Relevant information must be included in the training of participating professionals.

6. As the success of such programs depends on their quality, great attention must be focused on implementing proven national and international experiences, on training professionals to participate in the program, and on regular assessment of the program's impact.

Concerning *preventative health care*, it is most important to intervene prior to and directly following birth, and to counterbalance health disadvantages that develop in early childhood. As a series of epidemiological and health economics studies have shown in the past two decades,[47] pregnancy is a particularly sensitive period, which – by programming the fetal metabolism (*Wintour–Owens* [2006]) – has long-term consequences for the health and prospective diseases of the growing child and adult. Thus, it also influences life chances in a broad sense: school and labor market careers, income, life expectancy. The years preceding school enrollment, particularly the first three years of life, form a similarly important phase (*Almond–Currie* [2011a]). Children raised by less educated, low-income parents can amass serious health deficits at this age,[48] which can only

47 *Paul* [2010] provides an overview of factors that have a positive and negative impact on women during pregnancy and their consequences. An excellent summary of the relevant modern economic literature can be found in *Almond–Currie* [2011b], *Currie* [2011].

48 In Hungary, 17 % of Roma infants and 14-15 % of infants born to mothers who have not completed more than 8

be remedied at great cost – often only partially. Prevention holds great health and social potential. What must be done is clear; the Women Infants and Children (WIC) program[49] in the United States can serve as a model regarding the technical details. It is one of the largest and oldest programs of this kind, and its effect has been assessed on numerous occasions.[50]

WIC is a special supplemental nutrition program for pregnant and breastfeeding women, infants and children under 5 who live in poverty. The program offers participants supplemental nutritious foods (ensuring adequate intake of vitamins, minerals and proteins), detailed and individualized nutrition and preventative health counseling. Standardized screenings and referrals to other health services are meant to prevent neglect of diseases and developmental problems. It would clearly be a challenge to adapt a social program of this size and complexity for use in Hungary. However, there is a nearly unique resource which could enable such an undertaking: a network of visiting nurses which covers the entire country. Visiting nurses are currently the only state employees who know every child, who enter the homes of every family, even those who live in poverty, shut off from others. They are in possession of information that could form a solid basis for such a program. Visiting nurses have a direct, personal relationship with families from pregnancy to the end of early childhood. Hungary is thus in a uniquely good situation to implement a program like WIC.[51]

A further major factor in the school achievement gap is that, due to residential disadvantages and the selection mechanisms of the school system, Roma students are excluded from schools that offer high quality instruction. This phenomenon has various causes. One of the basic reasons – in our opinion – is that *the Hungarian school system did not undergo thorough modernization* in previous decades, and the majority of Hungarian schools have not succeeded in catching up[52] with international education trends (*Andor* [2005], *McKinsey Report* [2007], *McKinsey Report* [2010]). In today's world, strong basic skills have rapidly become vital to success in the labor market (*Murnane–Levy* [1996]), and disadvantaged social groups are the main victims of a relatively limited supply of good schools. To discuss possible solutions for this pervasive problem, even briefly, would exceed the scope of this article. However, we must emphasize that the Hungarian school system is in need of institutional reforms of the type and on the scale proposed by the

---

years of schooling weighed less than 2500g at birth at the beginning of the 1990s. For more, see the comparative data in footnote 15.

49 http://www.fns.usda.gov/wic. The program was established in 1972 and had 9.2 million participants (2.1 million pregnant women, 2.2 million newborns and 4.9 million children under 5) in the 2010 fiscal year. Its total budget was $ 6.7 billion in 2010. The rough Hungarian equivalent of this sum would be 50 billion HUF if we take population differences into consideration but ignore the difference in GDP per capita, and 12-13 billion HUF if we also take the difference in GDP into account.

50 http://www.fns.usda.gov/ora/MENU/Published/WIC/WIC.htm. *Bitler–Currie* [2005] offers a very thorough impact analysis.

51 Besides ensuring adequate funding, many other related issues must also first be resolved. See *Herczog* [2008] p. 38–39 and 42–43.

52 This is clearly corroborated by international studies (see *Csapó* [2012]).

Education and Children's Opportunity Round Table in a report entitled *Green Book for the Renewal of Hungarian Public Education* (*Zöld könyv* [2008]), which summarizes the results of two years of research.

The other cause is *school segregation.* There are great differences between Hungarian schools in terms of social and ethnic composition (*Csapó–Molnár–Kinyó* [2009], *Kertesi–Kézdi* [2009]). This has various reasons: differences in the social composition of residential areas, a high level of student mobility between school districts due to a system of free school choice (this process is highly selective by social status), selective admissions at better schools, local school policies that exacerbate segregation, and prejudice against Roma students. Anti-segregation policies must target the causes of segregation as well as sanction flagrant cases. The mechanisms responsible for segregation are extremely strong homeostatic processes, which is why only interventions that affect the system as a whole hold the promise of significant change.

However, despite the drawbacks of the school system, teacher distribution and the processes of segregation, implementation of the right *pedagogical innovations* could significantly improve the quality of instruction that disadvantaged students receive. Globally, the most successful school systems (*McKinsey Report* [2007], *McKinsey Report* [2010]) and successful experiments in other countries[53] have produced clear results. In Hungary, there are hardly any programs which build on these results to target the schools of disadvantaged students.[54]

The skill development and school careers of disadvantaged children – Roma children living in poverty among them – will largely depend on whether we prove to be capable of understanding, accepting, and using both Hungarian data and international experience. This is what we must build on to shape social policy in a way that uses available resources as efficiently as possible to help the children and their families.

53 For example, the Harlem Children's Zone program: http://www.hcz.org, KIPP schools: http://www.kipp.org and the *charter school* network: http://www.charterschoolcenter.org. Roland Fryer and his colleagues (www.edlabs.harvard.edu) have recently shown that the success of charter schools in the United States is due to the consistent implementation of a few extremely simple principles: frequent teacher feedback, the use of data to guide instruction, high-dosage tutoring, increased instructional time and high expectations (*Dobbie–Fryer* [2011b]). Applying these practices in traditional public schools also leads to good outcomes (*Fryer* [2012]).

54 There are individual examples that appear to be successful, such as the adaptation of the KIPP school methodology in Hejőkeresztúr (http://www.hejokeresztur.hu/index.php?option=com_content&view=art icle&id=161&Itemid=501) and the H2O school network (http:// h2oktatas.hu/hu/h2o-iskolak).

References

Akerlof, G. A.–Kranton, R. E. [2002]: Identity and schooling. Some lessons for the economics of education. Journal of Economic Literature, Vol. 40. No. 4. pp. 1167–1201.

Almond, D.–Currie, J. [2011a]: Human capital development before age five. In: *Ashenfelter, O.–Card, D.* (eds.): Handbook of Labor Economics. Vol. 4B. Elsevier, Amsterdam, pp. 1315–1486.

Almond, D.–Currie, J. [2011b]: Killing me softly: the fetal origins hypothesis. Journal of Economic Perspectives, Vol. 25. No. 3. pp. 153–172.

Amato, P. R. [2005]: The impact of family formation change on the cognitive, social, and emotional well-being of the next generation. The Future of Children, Vol. 15. No. 2. p. 75–96.

Ammermueller, A.–Pischke, J-S. [2009]: Peer effects in European primary schools. Journal of Labor Economics, Vol. 27. No. 3. pp. 315–348.

Andor Mihály [2005]: Lépéskényszer. Az extenzív fejlődés lehetőségeinek kimerülése az oktatásban. Iskolakultúra, Vol. 15. No. 3. pp. 57–70. http://epa.oszk.hu/00000/00011/00091/pdf/iskolakultura_EPA00011_2005_03_057-070.pdf.

Angrist, J D.–Dynarski, S. M.–Kane, T. J.–Pathak, P. A.–Walters, C. R. [2010]: Inputs and impacts in charter schools: KIPP Lynn. American Economic Review, Vol. 100. No. 2. pp. 1–5.

Antos, J. R.–Rosen, S. [1975]: Discrimination in the market for public school teachers. Journal of Econometrics, Vol. 3. No. 2. pp. 123–150.

Barajas, R. G.–Philipsen, N.–Brooks-Gunn, J. [2007]: Cognitive and emotional outcomes for children in poverty. In: Handbook of families and poverty. Sage, Thousand Oaks, CA, pp. 311–333.

Barker, D. J. P. [1998]: Mothers, babies and health in later life. Churchill Livingstone, Edinburgh.

Barro, Robert and Jong-Wha Lee, "A New Data Set of Educational Attainment in the World, 1950-2010." forthcoming, Journal of Development Economics.

Becker, G. S. [1981a]: The demand for children. In: 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: A treatise on the family. Harvard University Press, Cambridge, Mass. Enlarged edition: 1991, pp. 179–200.

Becker, G. S.–Tomes, N. [1986]: Human capital and the rise and fall of families. Journal of Labour Economics, Vol. 4. No. 3. pp. S1–S39.

Behrman, J. R.–Rosenzweig, M. R. [2004]: Returns to birthweight. Review of Economics and Statistics, Vol. 86. No. 2. pp. 586–601.

Bishop, J. H.–Bishop, M.–Gelbwasser, L.–Green, S.–Zuckerman, A. [2003]: Nerds and freaks: a theory of student culture and norms. Brookings Papers on Education Policy, pp. 141–213.

Bitler, M. P.–Currie, J. [2005]: Does WIC work? The effects of WIC on pregnancy and birth outcomes. Journal of Policy Analysis and Management, Vol. 24. No. 1. pp. 73–91.

Biztos Kezdet [2010]: A Biztos Kezdet program hatásvizsgálata. Kutatási összefoglaló. http://www.biztoskezdet.hu/uploads/attachments/BK_HV_bemeneti_meres_jav2010okt.pdf.

Black, S. E.–Devereux, P. J.–Salvanes, K. G. [2007]: From the cradle to the labor market. The effect of birth weight on adult outcomes. Quarterly Journal of Economics, Vol. 122. No. 1. pp. 409–439.

Black, S. E.–Devereux, P. J.–Løken, K. V.–Salvanes, K. G. [2012]: Care or cash? The effect of child care subsidies on student performance. NBER Working Paper No. 18086.

Boyd, D.–Lankford, H.–Loeb, S.–Wyckoff, J. [2003]: Analyzing the determinants of the matching of public school teachers to jobs: estimating compensating differentials in imperfect labor markets.

NBER Working Paper No. 9878.

Bradley, R. H. et al. [2000]: Measuring the HOME environments of children in early adolescence. Journal of Research on Adolescence, Vol. 10. No. 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, Vol. 148. No. 4. pp. 377–383.

Brooks-Gunn, J.–Berlin, L. J.–Fuligni, A. S. [2000]: Early childhood intervention programs. What about the family? In: *Shonkoff, J. P.–Meisels S. J.* (Eds.): Handbook of early childhood intervention, Cambridge University Press, Cambridge, pp. 549–577.

Brooks-Gunn, J.–Duncan, G. J. [1997]: The effects of poverty on children. The Future of Children, Vol. 7. No. 2. pp. 55–71.

Brooks-Gunn, J.–Markman, L. G. [2005]: The contribution of parenting to ethnic and racial gaps in school readiness. The Future of Children, Vol. 15. No. 1. pp. 139–168.

Card, D. [1999]: The causal effect of education on earnings. In: *Ashenfelter, O.–Card, D.* (Eds.): Handbook of Labor Economics, Vol. 3A. Elsevier, Amsterdam, pp. 1801–1863.

Case, A.–Fertig, A.–Paxson, C. [2005]: The lasting impact of childhood health and circumstance. Journal of Health Economics, Vol. 24. No. 2. pp. 365–389.

Case, A.–Lubotsky, D.–Paxson, C. [2002]: Economic status and health in childhood: the origins of the gradient. American Economic Review, Vol. 92. No. 5. pp. 1308–1334.

Chetty, R.–Friedman, J. N.–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.–Vigdor, J. [2005]: Who teaches whom? Race and the distribution of novice teachers. Economics of Education Review, Vol. 24. No. 4. pp. 377–392.

Conger, R. D.–Conger, K. J.–Elder, G. H.–Lorenz, F. O.–Simons, R. L.–Whitbeck, L. B. [1992]: A family process model of economic hardship and adjustment of early adolescent boys. Child Development, Vol. 63. No. 3. pp. 523–541.

Conger, R. D.–Conger, K. J.–Elder, G. H.–Lorenz, F. O.–Simons, R. L.–Whitbeck, L. B. [1993]: A family economic stress and adjustment of early adolescent girls. Development Psychology, Vol. 29. No. 2. pp. 206–219.

Cramer, J. C. [1995]: Racial and ethnic differences in birthweight: the role of income and financial assistance. Demography, Vol. 32. No. 2. pp. 231–247.

Crane, J. [1996]: Effects of home environment, SES, and maternal test scores on mathematics achievement. Journal of Educational Research, Vol. 89. No. 5. pp. 305–314.

Currie, J [2011]: Inequality at birth: some causes and consequences. American Economic Review, Vol. 101. No. 3. pp. 1–22.

Currie, J. [2005]: Health disparities and gaps in school readiness. The Future of Children, Vol. 15. No. 1. pp. 117–138.

Currie, J. [2009]: Healthy, wealthy, and wise. Journal of Economic Literature, Vol. 47. No. 1. pp. 87–122.

Currie, J.–Hyson, R. [1999]: Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. American Economic Review, Vol. 89. No. 2. pp. 245–250.

Currie, J.–Stabile, M. [2003]: Socioeconomic status and child health. Why is the relationship stronger for older children? American Economic Review, Vol. 93. No. 5. pp. 1813–1823.

Currie, J.–Stabile, M. [2009]: Mental health in childhood and human capital. In: *Gruber, J.* (Ed.): The problem of disadvantaged youth: an economic perspective. University of Chicago Press,

Chicago, pp. 115–148.

Csapó Benő [2007]: Hosszmetszeti felmérések iskolai kontextusban. Magyar Pedagógia, Vol. 107. No. 4. pp. 321–355.

Csapó Benő (Ed.) [2012]: Mérlegen a magyar iskola. Nemzeti Tanköyvkiadó, Budapest. http://www.ntk.hu/c/document_library/get_file?p_l_id=11131&folderId=23648832&name=DLFE-24695.pdf.

Csapó Benő–Molnár Gyöngyvér–Kinyó László [2009]: A magyar oktatási rendszer szelektivitása a nemzetközi összehasonlító vizsgálatok tükrében. Iskolakultúra, Vol. 19. No. 3–4. pp. 3–13.

Dahl, G. B.–Lochner, L. [2011]: The impact of family income on child achievement: evidence from the earned income tax credit. University of Western Ontario, CIBC Working Paper Series No. 3.

Dehejia, R. H.–Wahba, S. [2002]: Propensity score-matching methods for nonexperimental causal studies. Review of Economics and Statistics, Vol. 84. No. 1. pp. 151–161.

Dickinson, D. K.–Tabors, P. O. [2001]: Beginning literacy with language. Paul H. Brookes, Baltimore.

Dobbie, W.–Fryer, R. G. [2011a]: Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. American Economic Journal: Applied Economics, Vol. 3. No. 3. pp. 158–187.

Dobbie, W.–Fryer, R. G. [2011b]: Getting beneath the veil of effective schools: evidence from New York City. Manuscript, November, http://www.economics.harvard.edu/faculty/fryer/files/effective_schools.pdf.

Duncan, G. J.–Brooks-Gunn, J. (Eds.) [1997]: Consequences of growing up poor. Russel Sage Foundation, New York.

Duncan, G. J.–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.–Murnane, R. J. [2011b]: Introduction: the American dream, then and now. In: *Duncan, G. J.–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.–Preston, S. H. [1992]: Effects of early-life conditions on adult mortality: a review. Population Index, Vol. 58. No. 2. pp. 186–212.

Ferge Zsuzsa–Darvas Ágnes (Eds.) [2010]: Gyerekesélyek Magyarországon. Budapest.

Ferguson, R. F. [1998]: Can schools narrow the black-white test score gap? In: *Jencks, C.–Phillips, M.* (Eds.): The Black-White test score gap. Brookings Institution Press, Washington, D.C., pp. 318–374.

Fiszbein, A.–Schady, N. [2009]: Conditional cash transfers reducing present and future poverty. The World Bank, Washington.

FRA [2009]: European Union Minorities and Discrimination Survey. Data in Focus Report: The Roma. European Agency ofr Fundamental Rights.

Fryer, R. G. [2011]: Racial inequality in the 21th century: the declining significance of discrimination. In: *Ashenfelter, O.–Card, D.* (Eds.): Handbook of Labor Economics. Vol. 4B. Elsevier, Amsterdam, pp. 855–971.

Fryer, R. G. [2012]: Injecting successful charter school strategies into traditional public schools: early results from an experiment in Houston. Manuscript, January, http://www.economics.harvard.edu/faculty/fryer/files/charter_school_strategies.pdf.

Fryer, R. G.–Torelli, P. [2010]: An empirical analysis of „acting white”. Journal of Public Economics, Vol. 94. No. 5–6. pp. 380–396.

Furstenberg, F. F. [2011]: The challanges of finding causal links between family educational practices and schooling outcomes. In: *Duncan, G. J.–Murnane, R. J.* (Eds.) [2011a] pp. 465–482.

Gould, E. D.–Simhon, A. [2011]: Does quality time produce quality children? IZA Discussion Paper No. 5487.

Guo, G.–Harris, K. M. [2000]: The mechanisms mediating the effects of poverty on children’s intellectual development. Demography, Vol. 37. No. 4. pp. 431–447.

Guryan, J.–Hearst, E.–Kearney, M. [2008]: Parental education and parental time with children. Journal of Economic Perspectives, Vol. 22. No. 3. pp. 23–46.

Hack, M.–Klein, N. K. –Taylor, H. G. [1995]: Long-term developmental outcomes of low-birth weight infants. The Future of Children, Vol. 5. No. 1. pp. 176–196.

Hanushek, E. A.–Kain, J. F.–Markman, J. M.–Rivkin, S. G. [2003]: Does peer ability affects student achievement? Journal of Applied Econometrics, Vol. 18. No. 5. pp. 527– 544.

Hanushek, E. A.–Kain, J. F.–Rivkin, S. G. [2004]: Why public schools lose teachers? Journal of Human Resources, Vol. 39. No. 2. pp. 326–354.

Hanushek, E. A.–Kain, J. F.–Rivkin, S. G. [2009]: Brown v. Board of Education: the complex effects of school racial composition on achievement. Journal of Labor Economics, Vol. 27. No. 3. pp. 349–383.

Hart, B.–Risley, T. [1995]: Meaningful differences in the everyday experience of young American children. Brookes, Baltimore.

Havas Gábor [2004]: Halmozottan hátrányos helyzetű gyerekek és az óvoda. Iskolakultúra, Vol. 14. No. 4. pp. 3–16.

Haveman, R.–Wolfe, B. [1995]: The determinants of children’s attainments: a review of methods and findings. Journal of Economic Literature, Vol. 33. No. 4. pp. 1829–1878.

Heath, S. B. [1984]: Ways with words: language, life, and work in communities and class- rooms. Cambridge University Press.

Heckman, J. J. [2011]: The American family in Black & White: a post-racial strategy for improving skills to promote equality. Daedalus, Vol. 140. No. 2. pp. 70–89.

Heckman, J. J.–Stixrud, J.–Urzua, S. [2006]: The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics, Vol. 24. No. 3. pp. 411–482.

Herczog Mária [2008]: A kora gyermekkori fejlődés elősegítése. In: *Fazekas Károly– Köllő János–Varga Júlia* (Eds): Zöld könyv a magyar közoktatás megújításáért. Ecostat, Budapest, pp. 33–52.

Hoff, E. [2003]: The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. Child Development, Vol. 74. No. 5. pp. 1368–1378.

Hoff, E. [2006]: How social contexts support and shape language development. Developmental Review, Vol. 26. No. 1. pp. 55–88.

Hoff-Ginsberg, E. [1991]: Mother-child conversation in different social classes and communicative settings. Child Development, Vol. 62. No. 4. pp. 782–796.

Huttenlocher, J.–Vasilyeva, M.–Cymerman, E.–Levine, S. [2002]: Language input and child syntax. Cognitive Psychology, Vol. 45. Iss. 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, Vol. 27. No. 2. pp. 213–256.

Józsa Krisztián [2004]: Az első osztályos tanulók elemi alapkészségeinek fejlettsége. Egy longitudinális kutatás első mérési pontja. Iskolakultúra, Vol. 14. No. 11. pp. 3–16.

Kalil, A.–DeLeire, T. [2004]: Family investments in children's potential. Resources and parenting behaviors that promote success. Lawrence Erlbaum, Mahwah (NJ).

Kaushal, N.–Magnuson, K.–Waldfogel, J. [2011]: How is family income related to investments in children's learning? In: *Duncan, G. J.–Murnane, R. J.* (Eds.) [2011a] pp. 187–205.

Kertesi Gábor [2000]: A cigány foglalkoztatás leépülése és szerkezeti átalakulása 1984 és 1994 között. Munkatörténeti elemzés. Közgazdasági Szemle, Vol. 47. No. 5. pp. 406–443.

Kertesi Gábor [2005]: Roma foglalkoztatás az ezredfordulón. A rendszerváltás maradandó sokkja. Szociológiai Szemle, Vol. 15. No. 2. pp. 57–87.

Kertesi Gábor–Kézdi Gábor [2005]: Általános iskolai szegregáció. Parts I–II. Közgazdasági Szemle, Vol. 52. No. 4. pp. 317–355 and No. 5. pp. 462–479.

Kertesi Gábor–Kézdi Gábor [2007]: Children of the post-communist transition: age at the time of the parents' job loss and dropping out of secondary school. The Berkeley Electronic Journal of Economic Analysis & Policy, Vol. 7. No. 2 (Contributions), Article 8.

Kertesi Gábor–Kézdi Gábor [2009]: Általános iskolai szegregáció az ezredforduló után. Közgazdasági Szemle, Vol. 56. No. 11. pp. 959–1000.

Kertesi Gábor–Kézdi Gábor [2010]: Iskolázatlan szülők gyermekei és roma fiatalok a középiskolában. In: *Kolosi Tamás–Tóth István György* (Eds.): Társadalmi riport, 2010. Tárki, Budapest, pp. 371–407.

Kertesi Gábor–Kézdi Gábor [2011]: The Roma/non-Roma test score gap in Hungary. American Economic Review, Vol. 101. No. 3. pp. 519–525.

Kertesi Gábor–Kézdi Gábor [2012a]: Residential segregation, local policies, and school segregation in 100 Hungarian towns. Roma Educational Fund, Budapest, manuscript.

Kertesi Gábor–Kézdi Gábor [2012b]: A roma és nem roma tanulók teszteredményei közti különbségekről és e különbségek okairól. BWP, No. 5 http://www.econ.core.hu/file/download/bwp/bwp1205.pdf.

Kertesi Gábor–Varga Júlia [2005]: Foglalkoztatás és iskolázottság Magyarországon. Közgazdasági Szemle, Vol. 52. No. 7–8. pp. 633–662.

Kézdi Gábor–Köllő János–Varga Júlia [2008]: Az érettségit nem adó szakmunkásképzés válságtünetei. In: *Fazekas Károly–Köllő János* (Eds.): Munkaerőpiaci Tükör, 2008. MTA Közgazdaságtudományi Intézet–OFA, Budapest, pp. 87–136.

Kézdi Gábor–Surányi Éva [2008]: Egy sikeres iskolai integrációs program tapasztalatai. Educatio, Budapest.

Kézdi Gábor–Varga Júlia [2007]: Téveszmék és álterápiák. Figyelő, január 11.

Köllő János [1997]: A napi ingázás feltételei és a helyi munkanélküliség Magyarországon: számítások és szám példák. Esély, Vol. 8. No. 2. pp. 33–61.

Lareau, A. [2003]: Unequal childhoods. University of California Press. Berkeley.

Lee, V.–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, Vol. 2. No. 2. pp. S111–S131.

Lempers, J. D.–Clark-Lempers, D.–Simons, R. L. [1989]: Economic hardship, parenting, and distress in adolescence. Child Development, Vol. 60. No. 10. pp. 25–39.

Linver, M. R.–Brooks-Gunn, J.–Kohen, D. [2002]: Family processes as pathways from income to young children's development. Development Psychology, Vol. 38. No. 5. pp. 719–734.

Løken, K. V., Mogstad, M.–Wiswall, M. [2012]: What linear estimators miss: the effects of family income on child outcomes. American Economic Journal: Applied Economics, Vol. 4. No. 2. pp. 1–35.

Lőrik József [1999]: A gyermeki szókincs vizsgálata. In: *Földes Tamás* (Ed.): Diagnosztizálás, értékelés, minőségbiztosítás. „Fogjuk a kezed" Egyesület, Sopron.

Marmot, M.–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.

McKinsey-jelentés [2007]: Mi áll a világ legsikeresebb iskolai rendszerei teljesítményének hátterében? Eds.: *Barber, M.–Mourshed, M.* http://mek.oszk.hu/09500/09575/09575.pdf.

McKinsey Report [2010]: How the world's most improved school systems keep getting better. Eds.: *Barber, M.–Mourshed, M.* http://mckinseyonsociety.com/downloads/reports/Education/How-the-Worlds-Most-Improved-School-Systems-Keep-Getting-Better_Download-version_Final.pdf.

McLanahan, S. [2004]: Diverging destinies: how children are faring under the second demographic transition. Demography, Vol. 41. No. 4. pp. 607–627.

McLanahan, S.–Percheski, C. [2008]: Family structure and the reproduction of inequalities. Annual Review of Sociology, Vol. 34. pp. 257–276.

McLanahan, S.–Sandefur, G. [1994]: Growing up with a single parent. 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, Vol. 61. No. 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.

Medgyesi Márton [2007]: Az otthoni környezet felmérése az Életpálya-2006 vizsgálatban: HOME inventory scale. Manuscript.

Moffitt, T. E. et al. [2011]: A gradient of childhood self-control predicts health, wealth and public safety. Proceedings of the National Academy of Sciences of the United States of America, Vol. 108. No. 7, February 15. pp. 2693–2698.

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, Vol. 4. No. 2–3. pp. 259–270.

Mulligan, C. B. [1997]: Parental priorities and economic inequality. The University of Chicago Press, Chicago–London.

Murnane, R. J.–Levy, F. [1996]: Teaching the new basic skills. Principles for educating children to thrive in a changing economy. The Free Press, New York.

Neal, D. A.–Johnson, W. R. [1996]: The role of premarket factors in Black-White wage differences. Journal of Political Economy, Vol. 104. No. 5. pp. 869–895.

Neuman, S. B. [1996]: Children engaging in storybook reading. Early Childhood Research Quarterly, Vol. 11. No. 4. pp. 495–513.

Neuman, S. B. [1999]: Books make a difference. A study of access to literacy. Reading Research

Quarterly, Vol. 34. No. 3. pp. 286–311.

Neuman, S. B. [2006]: The knowledge gap. In: *Dickinson, D. K.–Neuman, S. B.* (eds.): Handbook of early literacy research. Vol 2. The Guilford Press, New York–London. pp. 29–40.

Neuman, S. B.–Celano, D. C. [2001]: Access to print in low-income and middle-income communities: an ecological study of four neighborhoods. Reading Research Quarterly, Vol. 36. No. 1. pp. 8–26.

Neuman, S. B.–Celano, D. C. [2004]: Save the libraries! Educational Leadership, Vol. 61. No. 6. pp. 82–85.

Neuman, S. B.–Celano, D. C.–Greco, A. N.–Shue, P. [2001]: Access for all. Closing the book gap for children in early education. International Reading Association, Newark, DE.

OECD [2010]: PISA 2009 results: overcoming social background – Equity in learning opportunities and outcomes. Vol. II. OECD, Paris.

OECD [2011]: Divided we stand. Why inequality keeps rising. OECD, Paris.

Oreopoulos, P.–Stabile, M.–Walld, R.–Roos, L. L. [2008]: Short-, medium-, and long-term consequences of poor infant health. Journal of Human Resources, Vol. 43. No. 1. pp. 88–138.

Palloni, A.–Milesi, C.–White, R. G.–Turner, A. [2009]: Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials. Social Science & Medicine, Vol. 68. No. 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.–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, Vol. 77. No. 4. pp. 924–953.

Ramey, G.–Ramey, V. A. [2010]: The rug rat race. Brookings Papers on Economic Activity, Vol. 41. No. 1. pp. 129–199.

Réger Zita [1990]: Utak a nyelvhez: nyelvi szocializáció – nyelvi hátrány. Akadémiai Kiadó, Budapest.

Réger Zita [1995]: Cigány gyerekek nyelvi problémái és iskolai esélyei. Iskolakultúra, Vol. 5. No.

24. pp. 102–106.

Reichman, N. E. [2005]: Low birth weigh and school readiness. The Future of Children, Vol. 15. No. 1. pp. 91–116.

Rivkin, S. G.–Hanushek, E. A.–Kain, J. F. [2005]: Teachers, schools, and academic achievement. Econometrica, Vol. 73. No. 2. pp. 417–458.

Rosenzweig, M. R.–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: Economic aspects of health. *Fuchs, V. R.* (Ed.): University of Chicago Press, NBER, pp. 53–92.

Rouse, C.–Brooks-Gunn, J.–McLanahan, S. (Eds.) [2005]: School readiness: racial and ethnic gaps. The Future of Children, Vol. 15. No. 1. pp. 3–195.

Sayer, L. C.–Gauthier, A. H.–Furstenberg, F. F. [2004]: Educational differences in parents' time with children: cross-national variations. Journal of Marriage and Family, Vol. 66. No. 5. pp. 1152–1169.

Scafidi, B.–Sjoquist, D. L.–Stinebricker, T. R. [2007]: Race, poverty, and teacher mobility.

Economics of Education Review, Vol. 26. No. 2. pp. 145–159.

Schacter, J.–Thum, Y. M. [2004]: Paying for high and low quality teaching. Economics of Education Review, Vol. 23. No. 4. pp. 411–430.

Schonkoff, J. P.–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.–Lawson, E. P. [1996]: Knowledge of storybooks as a predictor of young children's vocabulary. Journal of Educational Psychology, Vol. 88. No. 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, Vol. 13. No. 2. pp. 145–166.

Smith, J. P. [2009]: Reconstructing childhood health histories. Demography, Vol. 46. No. 2. pp. 387–403.

Strauss, J.–Thomas, D. [2008]: Health over the life course. In: *Schultz, T. P.– Strauss, J.* (Eds.): Handbook of development economics. Vol. 4. Elsevier, Amsterdam, pp. 3376–3474.

Sweeney, M. M. [2011]: Family structure instability and adolescent educational outcomes. In: *Duncan, G. J.–Murnane, R. J.* (Eds.) [2011a] pp. 229–252.

Todd, P. E.–Wolpin, K. I. [2007]: The production of cognitive achievement in children: home, school, and racial test score gaps. Journal of Human Capital, Vol. 1. No. 1. pp. 91–136.

Waldfogel, J. [2006]: What children need. Harvard University Press, Cambridge, Mass.

Ward, W. P. [1992]: Birth weight and economic growth. Women's living standards in the industrializing West. The University of Chicago Press, Chicago–London.

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.

Wintour, E. M.–Owens, J. A. (Eds.) [2006]: Early life origins of health and disease. Springer, New York.

Zöld könyv [2008]: Zöld könyv a magyar közoktatás megújításáért. Eds: *Fazekas Károly–Köllő János–Varga Júlia.* Ecostat, Budapest, http://mek.oszk.hu/08200/08222/08222.pdf.

Appendix

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

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

		Standardized test score average [a]		Proportion of students whose mother	
	Number of observations	Reading	Mathematics	has completed no more than 8 years of school	has completed higher education
Based on National Assessment of Basic Competencies data on 8th 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 [b]	10,022	-0.11	-0.05	0.21	0.20
Sample in the second wave [b]	9,300	-0.10	-0.04	0.21	0.20
The sample that forms the basis of our estimates [b]	9,056	-0.09	-0.03	0.20	0.20

[a] 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)
[b] The statistics drawn from the HLCS are weighted values (using the sampling weights)

A3. The basic data for 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	

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	-0.23	-1.047	-0.324
	(0.053)**	(0.055)**	(0.048)**	(0.050)**
Biological mother in household		0.05		-0.048
		(0.231)		(0.253)
Non-biological mother in household		-0.19		-0.218
		(0.240)		(0.266)
Biological father in household		0.01		-0.176
		(0.389)		(0.217)
Non-biological father in household		-0.03		-0.261
		(0.389)		(0.219)
Mother's education: grades 0-8		-0.67		-0.659
		(0.048)**		(0.050)**
Mother's education: vocational school		-0.57		-0.527
		(0.038)**		(0.042)**
Mother's education: higher education		-0.26		-0.223
		(0.033)**		(0.038)**
Father's education: grades 0-8		-0.62		-0.708
		(0.053)**		(0.061)**
Father's education: vocational school		-0.43		-0.54
		(0.040)**		(0.047)**
Father's education: high school diploma		-0.25		-0.265
		(0.039)**		(0.047)**
Mother employed		-0.02		-0.008
		(0.035)		(0.037)
Father employed		0.03		-0.007
		(0.041)		(0.042)
Proportion of years mother employed while child was age 0-14		-0.01		-0.007
		(0.044)		(0.050)
Proportion of years father employed while child was age 0-14		0.19		0.117
		(0.051)**		(0.057)*
Logarithm of family income		0.00		0.047
		(0.028)		(0.031)
Logarithm of household size		-0.05		-0.02
		(0.055)		(0.062)
Number of unemployed adults		-0.03		-0.02
		(0.018)		(0.019)
Size of apartment, m^2 per person		0.00		0.001
		(0.002)		(0.002)
Number of rooms per person		0.23		0.227
		(0.057)**		(0.065)**
Bathroom in apartment		0.14		0.133
		(0.062)*		(0.062)*
No money for food		-0.20		-0.153
		(0.050)**		(0.052)**
No money for heating		-0.08		-0.058
		(0.036)*		(0.037)
Received regularized child-rearing assistance		0.04		0
		(0.031)		(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^2	0.06	0.27	0.07	0.27

Robust standard errors clustered by school in parentheses
*Significant at 5% level; **Significant at 1% level

A5. The results of the linear Oaxaca-Blinder decompositions for *Figure 4* (dependent variables: test scores, independent variables: family background)

	Dependent variable	
	Reading test scores	Mathematics test scores
Raw gap	- 0.97**	- 1.04**
Gap caused by composition effect	- 0.73**	- 0.74**
Gap caused by parameters	- 0.23**	- 0.39**
Gap caused by interaction	- 0.01	0.09

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

A6. The basic data for 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	

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^2 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			-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^2	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

A8.1. Detailed OLS regression estimates for *Table 8* (dependent variables: probability of bedtime storytelling during kindergarten, independent variables: family background)

	Dependent variable							
	Seldom or never told bedtime stories (child's response)		Often told bedtime stories (child's response)		Seldom or never told bedtime stories (parent's response)		Often told bedtime stories (parent's response)	
Roma	0.23	0.05	-0.30	-0.02	0.15	0.05	-0.27	-0.03
	(0.022)**	(0.036)	(0.022)**	(0.040)	(0.019)**	(0.025)*	(0.019)**	(0.039)
Biological mother in household		-0.12		0.16		-0.01		0.31
		(0.162)		(0.235)		(0.081)		(0.279)
Non-biological mother in household		0.00		0.02		-0.02		0.29
		(0.169)		(0.242)		(0.087)		(0.289)
Biological father in household		-0.22		0.23		-0.08		0.05
		(0.145)		(0.267)		(0.070)		(0.251)
Non-biological father in household		-0.24		0.24		-0.08		0.05
		(0.145)		(0.267)		(0.070)		(0.252)
Mother's education: grades 0-8		0.11		-0.24		0.04		-0.27
		(0.028)**		(0.039)**		(0.018)*		(0.042)**
Mother's education: vocational school		0.06		-0.16		0.00		-0.21
		(0.021)**		(0.031)**		(0.011)		(0.036)**
Mother's education: higher education		0.03		-0.07		0.00		-0.14
		(0.016)*		(0.026)**		(0.008)		(0.029)**
Father's education: grades 0-8		0.06		-0.08		0.01		-0.13
		(0.034)		(0.046)		(0.019)		(0.046)**
Father's education: vocational school		0.01		-0.05		-0.01		-0.05
		(0.020)		(0.033)		(0.010)		(0.038)
Father's education: high school diploma		0.00		-0.02		-0.01		-0.01
		(0.019)		(0.033)		(0.010)		(0.036)
Mother employed		-0.01		-0.03		-0.01		-0.02
		(0.020)		(0.029)		(0.012)		(0.029)
Father employed		-0.02		0.01		0.00		-0.05
		(0.023)		(0.033)		(0.014)		(0.030)
Proportion of years mother employed while child was age 0-14		0.01		0.03		0.00		-0.03
		(0.028)		(0.038)		(0.016)		(0.036)
Proportion of years father employed while child was age 0-14		-0.03		0.06		-0.02		0.06
		(0.036)		(0.041)		(0.021)		(0.040)
Logarithm of family income		0.00		-0.01		-0.02		0.01
		(0.017)		(0.022)		(0.011)		(0.024)
Logarithm of household size		0.08		-0.10		0.03		-0.04
		(0.035)*		(0.045)*		(0.021)		(0.046)
Number of unemployed adults		0.00		0.00		0.00		0.00
		(0.010)		(0.015)		(0.006)		(0.014)
Apartment size m2 per person		0.00		0.00		0.00		0.00
		(0.001)		(0.001)		(0.000)		(0.001)
Number of rooms per person		-0.06		0.11		-0.01		0.04
		(0.030)		(0.046)*		(0.017)		(0.049)
Bathroom in apartment		-0.10		0.04		-0.08		0.01
		(0.039)**		(0.039)		(0.030)**		(0.037)
No money for food		0.04		-0.05		0.04		0.00
		(0.032)		(0.041)		(0.021)		(0.039)
No money for		0.02		-0.03		-0.01		-0.03

heating		(0.024)		(0.029)		(0.014)		(0.028)
Received regularized child-rearing assistance		0.02		-0.01		0.00		-0.05
		(0.019)		(0.026)		(0.012)		(0.025)
Free lunch in 8th grade		-0.02		0.02		0.01		-0.03
		(0.026)		(0.036)		(0.013)		(0.034)
Free textbooks in 8th grade		0.00		0.02		-0.01		0.02
		(0.016)		(0.022)		(0.009)		(0.024)
Mother's education - data missing		0.09		-0.19		0.03		-0.05
		(0.150)		(0.227)		(0.072)		(0.272)
Father's education - data missing		-0.22		0.18		-0.09		0.02
		(0.144)		(0.267)		(0.068)		(0.254)
Family income - data missing		-0.01		0.00		-0.01		-0.05
		(0.019)		(0.031)		(0.011)		(0.033)
Size of apartment - data missing		-0.01		-0.09		0.13		-0.10
		(0.080)		(0.080)		(0.069)		(0.068)
Number of rooms - data missing		-0.04		0.06		-0.10		-0.08
		(0.080)		(0.140)		(0.041)*		(0.158)
Bathroom - data missing		0.28		-0.23		0.04		-0.15
		(0.156)		(0.197)		(0.079)		(0.144)
Poverty indicator - data missing		0.03		-0.06		-0.02		-0.08
		(0.081)		(0.100)		(0.028)		(0.087)
Region: Central		0.07		-0.46		-0.03		0.47
		(0.128)		(0.599)		(0.049)		(0.128)**
Region: Central Transdanubia		-0.17		0.11		0.26		1.22
		(0.238)		(0.603)		(0.219)		(0.348)**
Region: Western Transdanubia		0.11		-0.12		0.43		0.79
		(0.248)		(0.598)		(0.288)		(0.222)**
Region: Southern Transdanubia		-0.06		0.10		0.25		0.98
		(0.179)		(0.608)		(0.214)		(0.374)**
Region: Northern Hungary		-0.04		-0.48		0.00		0.17
		(0.243)		(0.608)		(0.051)		(0.300)
Region: Northern Great Plain		0.02		-0.11		0.04		0.73
		(0.186)		(0.674)		(0.057)		(0.183)**
Budapest		-0.04		-0.13		0.03		-0.06
		(0.053)		(0.111)		(0.023)		(0.108)
County seat		-0.02		0.03		0.00		-0.02
		(0.039)		(0.057)		(0.022)		(0.065)
Other city		-0.06		-0.03		0.01		-0.02
		(0.041)		(0.055)		(0.018)		(0.058)
Remote settlement		0.00		-0.01		0.03		0.02
		(0.032)		(0.041)		(0.022)		(0.047)
Constant		0.44		0.66		0.24		-0.36
		(0.290)		(0.643)		(0.172)		(0.380)
Number of observations	9056	9056	9056	9056	9056	9056	9056	9056
R^2	0.03	0.48	0.03	0.50	0.04	0.47	0.02	0.52

Robust standard errors clustered by school in parentheses
*Significant at 5% level; **Significant at 1% level

A8.2. Detailed OLS regression estimates for *Table 8* (dependent variables: probability of hiking during adolescence, cognitive and emotional HOME index, independent variables: family background)

	Dependent variable					
	Seldom went hiking with parents (child's response)		Cognitive HOME index		Emotional HOME index	
Roma	0.31	0.01	-1.12	-0.08	-0.18	0.07
	(0.021)**	(0.038)	(0.051)**	(0.070)	(0.049)**	(0.075)
Biological mother in household		-0.07		-0.39		-0.36
		(0.188)		(0.305)		(0.350)
Non-biological mother in household		-0.04		-0.59		-0.47
		(0.195)		(0.330)		(0.350)
Biological father in household		0.07		0.46		0.54
		(0.207)		(0.337)		(0.572)
Non-biological father in household		0.08		0.40		0.54
		(0.208)		(0.340)		(0.573)
Mother's education: grades 0-8		0.23		-0.78		-0.15
		(0.039)**		(0.066)**		(0.069)*
Mother's education: vocational school		0.17		-0.54		-0.08
		(0.033)**		(0.051)**		(0.060)
Mother's education: higher education		0.08		-0.30		-0.03
		(0.027)**		(0.041)**		(0.050)
Father's education: grades 0-8		0.11		-0.44		-0.18
		(0.045)*		(0.072)**		(0.086)*
Father's education: vocational school		0.06		-0.24		-0.08
		(0.036)		(0.053)**		(0.067)
Father's education: high school diploma		0.01		-0.10		-0.06
		(0.035)		(0.049)*		(0.067)
Mother employed		-0.03		0.07		0.02
		(0.028)		(0.047)		(0.048)
Father employed		-0.01		0.05		-0.12
		(0.031)		(0.054)		(0.063)
Proportion of years mother employed while child was age 0-14		-0.01		0.03		0.03
		(0.034)		(0.062)		(0.064)
Proportion of years father employed while child was age 0-14		-0.01		0.17		0.14
		(0.041)		(0.074)*		(0.084)
Logarithm of family income		-0.03		0.01		-0.06
		(0.021)		(0.038)		(0.045)
Logarithm of household size		0.03		0.07		0.12
		(0.042)		(0.072)		(0.080)
Number of unemployed adults		0.00		0.05		-0.04
		(0.013)		(0.025)		(0.027)
Size of apartment, m2 per person		0.00		0.00		0.00
		(0.001)		(0.002)		(0.002)
Number of rooms per person		-0.08		0.30		0.12
		(0.045)		(0.073)**		(0.085)
Bathroom in apartment		-0.06		0.53		0.14
		(0.037)		(0.082)**		(0.082)
No money for food		0.07		-0.15		-0.05
		(0.036)		(0.071)*		(0.078)

No money for heating		0.00 (0.027)		-0.18 (0.053)**		-0.07 (0.055)
Received regularized child-rearing assistance		0.03 (0.024)		-0.19 (0.039)**		-0.04 (0.047)
Free lunch in 8th grade		0.00 (0.036)		-0.03 (0.056)		-0.01 (0.068)
Free textbooks in 8th grade		0.00 (0.022)		-0.02 (0.036)		-0.08 (0.045)
Mother's education - data missing		0.07 (0.178)		-1.18 (0.288)**		-0.28 (0.345)
Father's education - data missing		0.10 (0.209)		0.29 (0.342)		-0.14 (0.572)
Family income - data missing		-0.01 (0.028)		0.00 (0.048)		0.01 (0.056)
Size of apartment - data missing		0.00 (0.061)		-0.07 (0.141)		-0.09 (0.169)
Number of rooms - data missing		-0.08 (0.095)		0.14 (0.303)		0.28 (0.196)
Bathroom - data missing		0.16 (0.153)		-0.11 (0.415)		0.15 (0.259)
Poverty indicator - data missing		-0.10 (0.090)		-0.12 (0.135)		0.26 (0.192)
Region: Central		-0.09 (0.284)		0.24 (0.182)		0.97 (0.256)**
Region: Central Transdanubia		-0.38 (0.458)		1.07 (0.216)**		0.87 (0.486)
Region: Western Transdanubia		-0.03 (0.537)		1.29 (0.196)**		1.06 (0.531)*
Region: Southern Transdanubia		0.25 (0.523)		0.83 (0.246)**		1.42 (0.437)**
Region: Northern Hungary		0.11 (0.342)		0.08 (0.460)		0.24 (0.433)
Region: Northern Great Plain		0.12 (0.330)		0.18 (0.304)		0.48 (0.426)
Budapest		0.09 (0.081)		0.15 (0.165)		-0.36 (0.246)
County seat		-0.14 (0.067)*		0.10 (0.096)		-0.06 (0.120)
Other city		-0.02 (0.041)		0.06 (0.087)		-0.03 (0.109)
Remote settlement		0.06 (0.040)		-0.09 (0.064)		0.01 (0.089)
Constant		0.79 (0.470)		-1.09 (0.613)		-0.15 (0.760)
Number of observations	9056	9056	9056	9056	9056	9056
R^2	0.03	0.57	0.09	0.70	0.00	0.61

Robust standard errors clustered by school in parentheses
*Significant at 5% level; **Significant at 1% level

A8.3. Detailed OLS regression estimates for *Table 8* (dependent variables: few or no books, probability of an Internet connection at home, independent variables: family background)

	Dependent variable			
	No or very few books at home		Internet connection at home	
Roma	0.55	0.24	-0.44	-0.05
	(0.013)**	(0.040)**	0.00	(0.027)
Biological mother in household		0.06		-0.10
		(0.138)		(0.180)
Non-biological mother in household		0.02		-0.21
		(0.142)		(0.182)
Biological father in household		-0.06		0.34
		(0.180)		(0.169)*
Non-biological father in household		-0.06		0.31
		(0.180)		(0.169)
Mother's education: grades 0-8		0.15		-0.24
		(0.023)**		(0.036)**
Mother's education: vocational school		0.04		-0.18
		(0.015)**		(0.031)**
Mother's education: higher education		-0.01		-0.05
		(0.010)		(0.027)
Father's education: grades 0-8		0.08		-0.22
		(0.026)**		(0.038)**
Father's education: vocational school		-0.01		-0.16
		(0.013)		(0.032)**
Father's education: high school diploma		-0.02		-0.07
		(0.011)*		(0.032)*
Mother employed		0.00		0.01
		(0.018)		(0.024)
Father employed		-0.01		0.06
		(0.021)		(0.026)*
Proportion of years mother employed while child was age 0-14		-0.04		0.04
		(0.023)		(0.031)
Proportion of years father employed while child was age 0-14		-0.02		-0.01
		(0.029)		(0.034)
Logarithm of family income		-0.01		0.05
		(0.013)		(0.021)*
Logarithm of household size		-0.02		0.08
		(0.031)		(0.039)*
Number of unemployed adults		-0.01		0.02
		(0.009)		(0.012)
Size of apartment, m^2 per person		0.00		0.00
		(0.001)		(0.001)
Number of rooms per person		-0.04		0.18
		(0.025)		(0.038)**
Bathroom in apartment		-0.16		0.01
		(0.042)**		(0.025)
No money for food		0.08		0.02
		(0.033)*		(0.028)
No money for heating		0.03		-0.02
		(0.022)		(0.023)
Received regularized child-rearing assistance		0.01		-0.05
		(0.016)		(0.021)*

Free lunch in 8th grade		0.05		0.01
		(0.023)*		(0.030)
Free textbooks in 8th grade		-0.02		-0.03
		(0.013)		(0.020)
Mother's education - data missing		0.15		-0.30
		(0.125)		(0.171)
Father's education - data missing		-0.08		0.17
		(0.179)		(0.171)
Family income - data missing		0.00		0.01
		(0.017)		(0.028)
Size of apartment - data missing		0.10		-0.04
		(0.063)		(0.070)
Number of rooms - data missing		-0.10		-0.13
		(0.081)		(0.112)
Bathroom - data missing		-0.07		-0.12
		(0.136)		(0.090)
Poverty indicator - data missing		-0.07		-0.07
		(0.051)		(0.097)
Region: Central		-0.14		-0.02
		(0.167)		(0.099)
Region: Central Transdanubia		-0.75		0.18
		(0.245)**		(0.372)
Region: Western Transdanubia		-0.67		-0.02
		(0.221)**		(0.513)
Region: Southern Transdanubia		-0.67		-0.37
		(0.217)**		(0.434)
Region: Northern Hungary		-0.29		-0.37
		(0.224)		(0.210)
Region: Northern Great Plain		-0.29		-0.06
		(0.195)		(0.131)
Budapest		0.01		0.09
		(0.036)		(0.089)
County seat		-0.02		0.11
		(0.029)		(0.063)
Other city		-0.01		0.09
		(0.033)		(0.047)
Remote settlement		-0.01		-0.06
		(0.030)		(0.040)
Constant		0.85		-0.43
		(0.284)**		(0.359)
Number of observations	9056	9056	9056	9056
R^2	0.19	0.63	0.05	0.65

Robust standard errors clustered by school in parentheses
*Significant at 5% level; **Significant at 1% level

A9. The results of the linear Oaxaca-Blinder decompositions for *Figures 6* and *7* (dependent variables: the indicators of the home environment, independent variables: family background)

	Dependent variable	
	Seldom or never told bedtime stories (child's response)	Often told bedtime stories (child's response)
Raw gap	+ 0.23**	- 0.30**
Gap caused by composition effect	+ 0.17**	- 0.25**
Gap caused by parameters	+ 0.03	+ 0.01
Gap caused by interaction	+ 0.03	- 0.06
	Seldom or never told bedtime stories (child's response)	Often told bedtime stories (child's response)
Raw gap	+ 0.23**	- 0.30**
Gap caused by composition effect	+ 0.17**	- 0.25**
Gap caused by parameters	+ 0.03	+ 0.01
Gap caused by interaction	+ 0.03	- 0.06
	Seldom or never told bedtime stories (parent's response)	Often told bedtime stories (parent's response)
Raw gap	+ 0.15**	- 0.27**
Gap caused by composition effect	+ 0.08**	- 0.24**
Gap caused by parameters	+ 0.08	- 0.00
Gap caused by interaction	- 0.01	- 0.03
	Seldom went hiking with parents (child's response)	
Raw gap	+ 0.31**	
Gap caused by composition effect	+ 0.30**	
Gap caused by parameters	- 0.00	
Gap caused by interaction	+ 0.01	
	Cognitive HOME index	Emotional HOME index
Raw gap	- 1.11**	- 0.18**
Gap caused by composition effect	- 1.00**	- 0.32**
Gap caused by parameters	- 0.16	+ 0.07
Gap caused by interaction	+ 0.05	+ 0.07
	No or very few books at home	Internet connection at home
Raw gap	+ 0.55**	- 0.43**
Gap caused by composition effect	+ 0.28**	- 0.43**
Gap caused by parameters	+ 0.08*	- 0.11*
Gap caused by interaction	+ 0.19**	+ 0.11*

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

A10. Detailed OLS regression estimates for *Table 10* (dependent variable: probability of being in a class highly segregated by ability, 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	0.143	0.207	0.123
	(0.028)**	(0.029)**	(0.029)**	(0.029)**
Biological mother in household		-0.074		-0.077
		(0.112)		(0.114)
Non-biological mother in household		-0.061		-0.083
		(0.118)		(0.121)
Biological father in household		0.066		0.096
		(0.151)		(0.163)
Non-biological father in household		0.093		0.118
		(0.151)		(0.163)
Mother's education: grades 0-8		0.166		0.095
		(0.019)**		(0.021)**
Mother's education: vocational school		0.081		0.036
		(0.014)**		(0.015)*
Mother's education: higher education		0.039		0.018
		(0.010)**		(0.010)
Father's education: grades 0-8		0.095		0.054
		(0.023)**		(0.023)*
Father's education: vocational school		0.042		0.019
		(0.014)**		(0.014)
Father's education: high school diploma		0.002		-0.006
		(0.012)		(0.012)
Mother employed		-0.016		-0.012
		(0.016)		(0.016)
Father employed		-0.038		-0.035
		(0.019)*		(0.018)
Proportion of years mother employed while child was age 0-14		-0.020		-0.017
		(0.021)		(0.021)
Proportion of years father employed while child was age 0-14		-0.008		0.003
		(0.027)		(0.026)
Logarithm of family income		-0.012		-0.008
		(0.011)		(0.011)
Logarithm of household size		0.008		0.015
		(0.025)		(0.025)
Number of unemployed adults		-0.001		0.003
		(0.009)		(0.009)
Size of apartment, m2 per person		0.000		0.001
		(0.001)		(0.001)
Number of rooms per person		-0.074		-0.046
		(0.024)**		(0.024)
Bathroom in apartment		-0.105		-0.077
		(0.031)**		(0.031)*
No money for food		0.022		0.011
		(0.024)		(0.023)
No money for heating		0.029		0.021
		(0.017)		(0.017)
Received regularized child-rearing assistance		-0.021		-0.031
		(0.015)		(0.015)*

Free lunch in 8th grade	0.055		0.050
	(0.024)*		(0.024)*
Free textbooks in 8th grade	0.031		0.032
	(0.012)**		(0.012)**
Mother's education - data missing	-0.024		-0.084
	(0.105)		(0.108)
Father's education - data missing	0.053		0.070
	(0.150)		(0.163)
Family income - data missing	-0.011		-0.008
	(0.014)		(0.014)
Size of apartment - data missing	0.118		0.100
	(0.056)*		(0.057)
Number of rooms - data missing	-0.093		-0.106
	(0.052)		(0.055)
Bathroom - data missing	0.026		0.005
	(0.095)		(0.095)
Poverty indicator - data missing	0.019		0.021
	(0.055)		(0.056)
Region: Central	0.051		0.063
	(0.030)		(0.030)*
Region: Central Transdanubia	0.011		0.024
	(0.029)		(0.029)
Region: Western Transdanubia	-0.049		-0.029
	(0.027)		(0.026)
Region: Southern Transdanubia	0.069		0.074
	(0.035)*		(0.034)*
Region: Northern Hungary	0.107		0.115
	(0.031)**		(0.031)**
Region: Northern Great Plain	0.101		0.095
	(0.029)**		(0.029)**
Budapest	-0.101		-0.081
	(0.030)**		(0.030)**
County seat	-0.084		-0.068
	(0.020)**		(0.020)**
Other city	-0.041		-0.033
	(0.020)*		(0.020)
Remote settlement	0.016		0.016
	(0.023)		(0.023)
Seldom or never told bedtime stories (child's response)		0.043	0.031
		(0.022)*	(0.020)
Often told bedtime stories (child's response)		0.003	0.001
		(0.012)	(0.012)
Seldom or never told bedtime stories (parent's response)		0.045	0.032
		(0.033)	(0.032)
Often told bedtime stories (parent's response)		-0.018	-0.009
		(0.011)	(0.011)
Seldom went hiking with parents (child's response)		0.039	0.023
		(0.011)**	(0.011)*
Cognitive HOME index		-0.052	-0.029
		(0.007)**	(0.007)**
Emotional HOME index		0.006	0.006
		(0.006)	(0.007)
Storytelling variable missing		-0.019	-0.008
		(0.022)	(0.021)
Cognitive HOME variable missing		0.080	0.057
		(0.056)	(0.052)

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^2	0.07	0.18	0.16	0.20

Robust standard errors clustered by school in parentheses
*Significant at 5% level; **Significant at 1% level

A11. The results of the linear Oaxaca-Blinder decompositions for *Figure 9* (dependent variable: probability of being in a class highly segregated by ability, independent variables: family background and family background combined with the home environment)

Dependent variable: probability of being in a class highly segregated by ability	Explanatory variables	
	Family background variables	Family background and home environment variables
Raw gap	+ 0.40**	+ 0.40**
Gap caused by composition effect	+ 0.26**	+ 0.28**
Gap caused by parameters	+ 0.11*	+ 0.10*
Gap caused by interaction	+ 0.03	+ 0.02

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