Compensatory advantage and inequality in educational aspirations*

L. Flóra Drucker†

July 21, 2023

Draft

Click here for the latest version.

Abstract

Using two administrative and survey-based datasets from Hungary, I look at how children from different socioeconomic backgrounds update their educational aspirations in response to having to repeat a grade late in primary school. I find that grade retention is detrimental to aspirations and later secondary school track choice, and on average, it affects children of lower socioeconomic backgrounds more adversely. The average effect masks heterogeneities by the reasons for repeating: those children who are likely to repeat seventh grade due to poor mathematics performance do not change their aspirations after retention, regardless of their socioeconomic background. However, they are less likely to attend a secondary school track that provides access to tertiary education. As we move towards higher performers in mathematics – and consequently, more heterogeneous reasons for repeating –, retention results in a larger drop in aspirations and the probability of a secondary track ending with a high school diploma. In both outcomes, high socioeconomic status largely and, in some cases, entirely offsets the adverse effects.

*I am grateful to Dániel Horn for supervision and fruitful discussions. I also thank Alexander Cappelen, Rebeka Éva Cook, Attila Gáspár, Hedvig Horváth, Tamás Keller, Mats Köster, Botond Kőszegi and Tímea Laura Molnár for discussions and helpful suggestions. Participants of conferences and seminars also gave valuable comments, which I am thankful for. I am indebted to the Databank of the Centre for Economic and Regional Studies for providing access to the databases.

†Institute of Economics, Centre for Economic and Regional Studies, Budapest

1 Introduction

There is a high correlation between parents and their children's education levels. In 2012, 52% of 25-32 year-olds had the same education level as their parents in OECD countries (OECD, 2015). Parental education influences children's education outcomes through several channels. First, parental background affects children's cognitive and non-cognitive skills at an early age. Parental investment in early childhood is crucial (see a review in Heckman and Mosso, 2014). Children from a high socioeconomic background are more intelligent, more altruistic, less risk-seeking, and more patient already at ages 7-9 (Falk et al., 2019). Besides entering school with different skills, children from different socioeconomic backgrounds also differ in their resources to cope with hardships during their educational careers.

A strand of the sociology literature explores a specific channel of low intergenerational educational mobility: the compensatory advantage of high socioeconomic status. After an adverse school event – like failing a subject or a wrong school choice – children from high socioeconomic backgrounds have much more resources to compensate for the negative shock than children from lower socioeconomic status, and consequently, they recover faster. This compensatory advantage reinforces the initial differences in educational outcomes by socioeconomic status (Bernardi and Cebolla-Boado, 2013; Bernardi, 2014; Bernardi and Grätz, 2015; Bernardi and Triventi, 2018; Bernardi and Valdés, 2021).

I address the compensatory advantage channel by looking at how the educational aspirations of children from different socioeconomic backgrounds change after having to repeat a grade at the end of their primary school career. As more difficult subjects come in higher grades of primary school, students who had difficulties earlier might have an even harder time in these higher grades. The original aim of grade retention is to give a second chance to students who failed one or more subjects for the first time. Therefore, it could decrease the inequality in students' performance and help worse students achieve higher education levels than if they were promoted to the next grade without the necessary qualifications. On the other hand, grade retention is often associated with a stigma, which makes it more difficult for the retained student to catch up. These educational hardships might cause a decline in lower-socioeconomic-status children' dream education level, while children of higher socioeconomic backgrounds might stay focused on the education level they initially wanted to achieve.

Aspirations act as reference points that induce motivation through loss aversion (Heath et al., 1999; Page et al., 2007). One's social environment highly influences these reference points, and so

aspirations may reinforce economic inequalities (Genicot and Ray, 2017, 2020). If aspirations are crucial in children's achievements, it is worth studying how adverse events shape them during their educational careers. There is little evidence on the effect of grade retention on aspirations. Hughes et al. (2013) find that parental expectations about their child's highest education level decrease after retention in the first grade of primary school. Decreased parental expectations then play a role in the negative effect of retention on the third-grade performance of children. Cham et al. (2015) do not find any effect of retention in primary grades on the motivation to finish secondary school in ninth grade. However, they find that retained students value a high school diploma more in ninth grade and feel that their teachers and peers are more likely to expect them to graduate.

To my best knowledge, this paper is the first to estimate the compensatory advantage of high socioeconomic status in educational aspirations after grade retention. Using administrative and survey data from Hungary, I first estimate how children update their educational aspirations from the sixth to the eighth grade of primary school, conditional on whether they had to repeat the seventh grade and their socioeconomic status. I proxy socioeconomic status with parental education and measure aspirations by the years of education children want to achieve. Then, I look at how changes in aspirations translate to secondary school track choices.

I find that both aspirations and the subsequent probability of enrolling in a secondary school track that gives access to tertiary education decrease if children have to repeat the seventh grade. The association of retention with aspirations and secondary school tracks is heterogeneous across parental education. After controlling for various factors that affect retention, children of low-educated parents with average aspirations in sixth grade reduce their desired education level by almost ten months (0.81 years). In contrast, for children of high-educated parents, this decrease is only about three months (0.23 years). The repeater–non-repeater gap in eighth-grade aspirations is highest for children of low-educated parents with high aspirations in sixth grade. On the other hand, the gap is constant across initial aspiration levels for children of high-educated parents. These differences then translate to differences in the chance of being in a secondary school in tenth grade that ends with a high school diploma. Furthermore, in line with the findings of Andrew (2014) and Contini and Salza (2022), retention also increases the probability of children missing from the tenth-grade sample – a proxy for dropping out of school before the tenth grade –, and high-educated parents also mitigate this increased chance.

In Hungary, students can be retained if they fail at least one subject at school or have been absent for a significant amount of time and cannot pass an equivalence test. To address the heterogeneity in the reasons for repeating, I first split the sample by the sixth-grade performance of children in

mathematics – a subject in which their knowledge can be fairly objectively assessed. For the worst-performing students, who are most likely to repeat the seventh grade because of poor mathematics performance, retention does not affect aspirations. However, among those who escape retention in this group, children of high-educated parents can increase their aspirations significantly more than those of low-educated parents. As we move toward higher performers in mathematics – towards groups with heterogeneous reasons for retention –, repeaters decrease their aspirations significantly by eighth grade. High-educated parents largely mitigate the decrease in aspirations in these groups. Repeaters in all groups are then less likely to end up in a secondary school that gives access to tertiary education. However, high-educated parents offset this decreased chance in the higher-performing group.

Finally, I look at the results of the same regressions in a smaller administrative database, in which I can control for the second main reason for repeating: illnesses during the school year. The patterns are mostly similar to those without controlling for the health outcomes. However, the number of observations – and consequently, the number of repeaters – is much lower in this database, so many coefficients become insignificant.

2 Data

I use two datasets for this paper. For most of the descriptive statistics and the baseline regressions, I use the National Assessment of Basic Competencies (National ABC) database that contains administrative data of all Hungarian students in the sixth, eighth, and tenth grades between 2008 and 2017. The dataset also includes data from a voluntary survey that children fill in at home. To learn more about the reasons behind grade retention, I use an administrative dataset, the Admin3 dataset, that covers half of the Hungarian population of ages 5-74 between 2003 and 2017. This dataset contains demographic and labor market data and healthcare-related variables and can also be linked to the National ABC dataset.

2.1 National Assessment of Basic Competencies data

The first data source I use is the National ABC database which covers the period between 2008 and 2017. The National ABC is a standardized mathematics and reading comprehension test that all Hungarian students take in the sixth, eighth, and tenth grades of public education. For children in the sample, compulsory education started at age 6 or 7 (depending on the birth month of the student), so the tests are taken by 12/13, 14/15 and 16/17 year-olds. All students must write the test, except some students with special education needs. The test is centralized, administered by

the Education Authority, and measures students' problem-solving skills in mathematics and reading comprehension. Students take the test in their school at the end of May. Students have a unique identifier assigned by the educational authority, so we can link the same student's sixth-, eighth-, and tenth-grade data. The National ABC database contains the standardized test scores and various background characteristics of students and their families from a background questionnaire.

There are three background questionnaires: on the student, institution, and if a school has multiple branches, branch level. Students complete the student questionnaire on paper at home with the help of their parents. Completing the questionnaire is voluntary. It contains 47 questions covering various topics regarding the student's academic progress (last year's GPA, last midterm GPA and marks from main subjects, number of years in kindergarten, grade retention in different phases of the educational career, educational aspirations, how much they like specific subjects, extracurricular activities), family background (status on regular child protection allowance, subsidized meals, family members living with the student, parents' age, education level, and labor market status), household characteristics (size of household, age composition, number of rooms, books, bathrooms, computers, internet access, etc.), family activities, and the student's perception of the wealth of the family compared to neighbors.

2.2 Administrative database

The second database I use is an administrative dataset (Admin3) that contains monthly data from half of the Hungarian population of ages 5-74 between 2003 and 2017 (see a detailed description in Sebők, 2019). The main file of Admin3 contains demographic data, labor market status, income, job characteristics, social transfers, and education status of each person in the dataset. An individual ID then links this dataset to administrative health care data provided by the National Health Insurance Fund (*Nemzeti Egészségbiztosítási Alapkezelő* - NEAK), and administrative and survey-based educational data from the National ABC database.

To better predict retention, I use the administrative health care data from Admin3. This dataset contains monthly data of each insured person about the number of visits to the general practitioner, costs of outpatient care covered by social security, costs paid by the insured on medication, cost of purchased medication covered by social security, costs of inpatient care covered by NEAK, and whether the person is eligible for prescription exemption.

2.3 Definition of variables and the sample

2.3.1 Main variables

The treatment variable is a dummy indicating retention in the seventh grade of primary school. First, I define retention between the sixth and eighth grades as more than two years passing between a child's first sixth- and eighth-grade tests. If there are more instances of the same child in the sixth-grade database, I code this child as someone who repeated the sixth grade. From the number of sixth-grade observations and the years passed between the first sixth- and eighth-grade occurrences, I can deduce if a child repeated the seventh grade.

The first outcome variable is the educational aspiration of the child in the eighth grade. In the questionnaire, children choose the education level they want to achieve from a list of qualifications from primary school to a doctoral degree. I create two variables from the answers. The first is educational aspirations in three categories: at most vocational qualification, high school diploma or a post-secondary non-tertiary qualification, and college or above. For the second variable, I assign years of education to each qualification level in the questionnaire: primary school is eight years, a vocational degree is 11 years, and so on (see the construction of the variable in more detail in Appendix Section A.7.1).

The second outcome variable is an indicator of the child being in a secondary school in the tenth grade that gives access to tertiary education. In Hungary, these were the academic and technical secondary schools in the sample period. Children could also attend a vocational secondary school, but that did not end with a high school diploma, an essential requirement for admission to higher education. The National ABC database includes administrative data about the type of institution the child attends in a particular grade. I use this information to create a dummy variable that takes the value 1 if the child is in an academic or technical secondary school and 0 if they are in a vocational secondary school in tenth grade.

I am interested in the compensatory advantage of high socioeconomic status in the effects of grade retention. Unfortunately, neither the National ABC nor the Admin3 database includes parental occupation or income data, which are essential elements of socioeconomic status besides education (see, e.g., Ganzeboom et al., 1992). However, the National ABC student questionnaire asks about the highest education level of both parents in fine categories: from unfinished primary education to a university degree. Therefore, I proxy socioeconomic status with parental education (and use the term parental education henceforth). I create a categorical variable: low-educated parents mean neither of the parents has a high school diploma, while high-educated parents mean

at least one parent has a high school diploma. I chose this distinction based on Falk et al. (2019).[1] Also, since there are very few repeaters among children of parents with tertiary education, drawing the line at a lower parental education level leaves enough repeaters in both parental education categories.

2.3.2 Sample

The starting sample consists of all children for whom I could link sixth-grade and eighth-grade observations. I restrict the sample to those children who were not retained until the sixth grade because this way, the treatment is the first big negative shock for everyone in their school career. The student questionnaire asks whether the child had to repeat a grade once or multiple times in different parts of their school career. I use this variable to exclude those who were already retained until the sixth grade.[2]

Since children take the National ABC test in May, when answering the questionnaire, they already suspect if they will be retained. Therefore, subsequent retention may affect sixth-grade aspirations. Due to this potential effect on pre-treatment variables, I also exclude those retained in sixth grade.

Finally, I exclude children in an academic secondary school in either the sixth or the eighth grade. This is around ten percent of the sample. These are very high-skilled children, mostly of high socioeconomic status, who were selected early into competitive eight- and six-year academic secondary schools (see Horn, 2013). The remaining 90 percent of the children spend all eight years in a primary school, so their aspirations are more heterogeneous, and they have yet to choose a secondary school track.

3 Grade retention rules in Hungary

Grade retention in Hungary is regulated by the 2011 Act on National Public Education.[3] In the first grade of primary school, parents can request retaining their child even if the child fulfilled all educational requirements. The school principal has to approve the request. Grade retention can stem from two sources from the second grade and above. First, if the child receives an insufficient

[1] They also use information about family income which, unfortunately, I do not have.

[2] I explain this variable in more detail in Appendix Section A.7. Children of low-educated parents already have a higher chance to repeat in lower grades of primary school than children of high-educated parents (7.5 percent vs. 1.5 percent were retained in the first four grades in my sample). This means that the children of low-educated parents in the estimation sample were more able to meet school requirements and catch up with their peers.

[3] See `https://net.jogtar.hu/jogszabaly?docid=a1100190.tv`

mark in at least one subject, they have to take a re-take exam in those subjects at the end of the summer. If they fail, they have to repeat the entire grade. Teachers have a say in deciding whether to fail someone in a subject or make them do some extra coursework to pass the subject without the re-take exam. The second source is absenteeism. If the child was absent from at least 250 classes, they have to take an equivalency test in the subjects they could not receive a final grade due to the absence. They have to pass all tests to proceed to the next grade. If they have at least 250 unjustified absences – not justified by either the parents or a doctor –, they have to repeat the entire grade without the possibility of an equivalency test. Retention is also automatic if the child receives an insufficient grade from at least three subjects.

Another rare reason for retention is if the family spends one or more years abroad. In this case, the children have to prove they learned everything their peers learned during the school year by passing a grading exam. Often the skills and knowledge they learn abroad are very different from what the Hungarian school system requires, and they cannot pass the grading exam, leading to retention (see the experiences of returning children in Árendás et al., 2022). This type of retention is similarly stigmatized to the one due to bad performance at school, so in the end, these children also face a negative shock. Migration is most pronounced among families with higher-educated parents who speak foreign languages well, although among the poorer, Roma families, also many migrate or flee to other countries in the hope of better living conditions (Árendás et al., 2022).

In practice, grade retention in Hungary is quite rare. Usually, two to three percent of students are retained in the higher grades of primary school, while retention in the final grade is even rarer: less than one percent in most years.

4 Descriptive statistics

4.1 Baseline aspirations

Table 1 shows the educational aspirations in sixth grade by parental education. Children with low-educated parents are significantly more likely not to aspire higher than a vocational education already in sixth grade than children of high-educated parents (24.91% vs. 4.12%). They are also significantly less likely to aspire for tertiary education (24.51% vs. 66.53%). While the highest share of low-educated parents' children (50.58%) aspire for a high-school diploma or a post-secondary non-tertiary qualification, the most desirable education level for high-educated parents' children is tertiary.

Figure 1 shows that the difference in aspirations between high- and low-educated parents' chil-

Aspirations in sixth grade	Low ed.	High ed.	Total
At most vocational	24.91	4.12	12.02
High school diploma or post-secondary non-tertiary	50.58	29.35	37.41
College or university	24.51	66.53	50.57
Observations	165088	269844	436014

Table 1: Sixth-grade educational aspirations by parental education

Note: Column percentages.

dren is there at every skill level measured by mathematics test scores.[4] This finding is consistent with the literature on the associations of socioeconomic status and educational aspirations (see, e.g., Akerlof and Kranton, 2002; Guyon and Huillery, 2020; Agasisti and Maragkou, 2023).

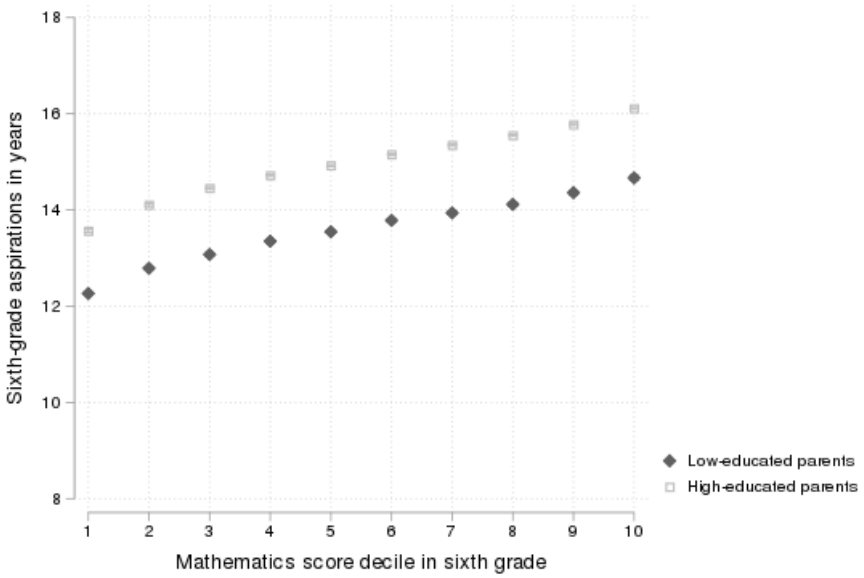

Figure 1: Sixth-grade educational aspirations by parental education and mathematics test scores

Note: Aspiration levels are converted to years of education. Spikes show 95 percent confidence intervals (though the intervals are so narrow that the spikes are virtually invisible).

4.1.1 Aspirations and secondary school tracks

Figure 2 shows that sixth-grade aspirations are good predictors of children's secondary track choice. Seventy percent of children who aspire to vocational education in sixth grade end up in a vocational secondary school, and only around five percent go to an academic secondary school. Ninety-seven

[4] Appendix Figure A1 shows the share of children with high-educated parents in each decile.

percent of those who desire to go to college or university in sixth grade attend a school in tenth grade that allows access to tertiary education.

The link between aspirations and secondary school tracks is even stronger if we look at eighth-grade aspirations (see Appendix Figure A2). We have to note that students take the National ABC at the end of May, while eighth-graders are notified about their admission to secondary schools at the end of April, so when stating their aspirations, most of them already know which secondary schools they will attend. Because of this, we cannot treat eighth-grade aspirations and tenth-grade secondary school outcomes completely separately, as they are highly correlated. However, the high correlation also validates the aspirations measure, strengthening the interpretation that aspirations reflect children's preferences about their future education.

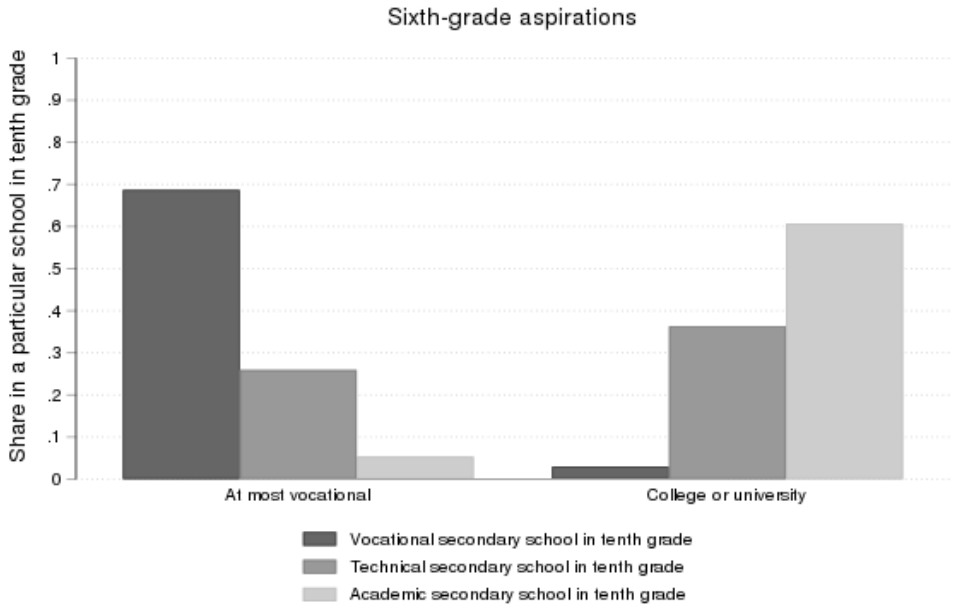

Figure 2: Share of students in secondary school tracks by sixth-grade aspirations

Note: The three bars on the left show the share of students in different secondary school tracks who wanted to achieve at most a vocational certificate in sixth grade. The three bars on the right show the same shares among those who aspired for tertiary education in sixth grade.

4.2 Retention

Table 2 shows the probability of children being retained in seventh grade by parental education, conditional on not having been retained until seventh grade. The number of observations is higher here than in Table 1 because repeater status is constructed from administrative data, while aspirations are self-reported. It is clear from the table that children of low-educated parents are more

likely to get retained: 1.28 percent of them repeat the seventh grade, compared to 0.37 percent of children with high-educated parents.

Repeated seventh grade	Low ed.	High ed.	Total
Did not repeat	98.72	99.63	99.28
Repeated	1.28	0.37	0.72
Observations	189880	308639	501315

Table 2: Repeating seventh grade grade by parental education

Note: Column percentages.

Figure 3 shows that, similarly to sixth-grade aspirations, repeating probabilities of children with low-educated parents are also higher at each mathematics test score decile.

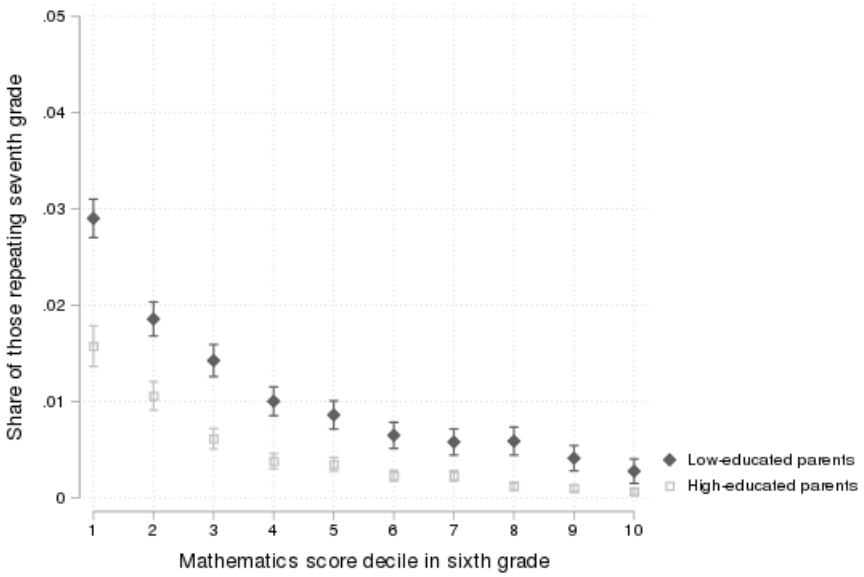

Figure 3: Share repeating seventh grade by mathematics test scores

Note: Spikes show 95 percent confidence intervals.

4.3 Change in aspirations after repeating

Figure 4 shows the average change in the aspired years of education for repeaters and non-repeaters by parental education. Those children who did not have to repeat seventh grade update their aspirations similarly, regardless of parental education. In both groups, children increase their aspired years of education by about 2 months (0.15-0.16 years) from sixth to eighth grade. However, there is a large difference between repeaters: retained children of low-educated parents decrease their

aspirations by 2 months (0.16 years), while there is no significant change in aspirations for retained children of high-educated parents.

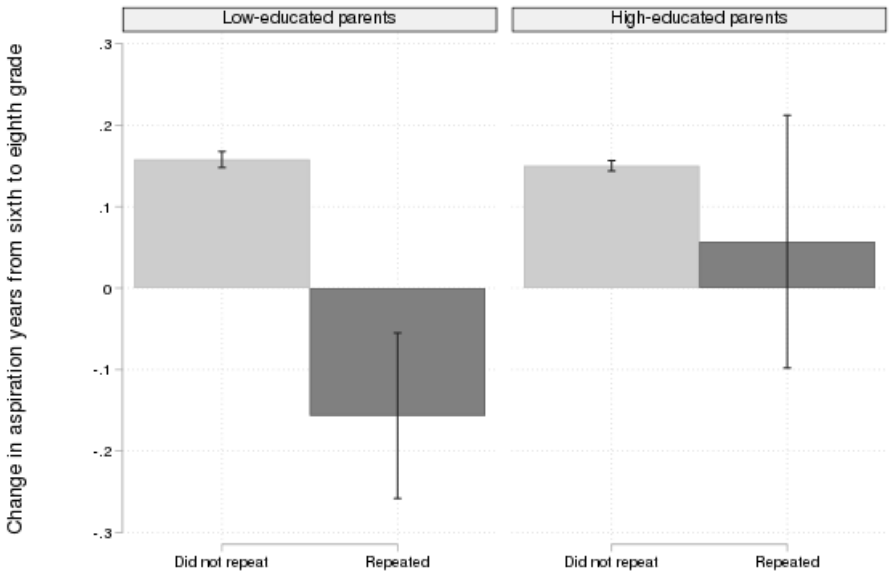

Figure 4: Average change in aspired years of education by repeater status and parental education

Note: Educational aspirations are presented in years. The spikes show 95 percent confidence intervals.

Figure 5 shows how educational aspirations change from sixth to eighth grade by parental education and repeater status at each level of initial aspirations. The figure shows desired education levels in years of education. While the aspiration changes of non-repeaters follow a similar trend across parental education, there are considerable differences between repeaters in the two groups. For children with low-educated parents, the aspiration gap between repeaters and non-repeaters increases with initial aspirations. In contrast, it stays pretty constant for children with high-educated parents, at around a year.

5 Empirical analysis

5.1 Predicting retention

The first column of Table 3 shows the raw difference between the repeating probabilities of high- and low-educated parents' children. Many factors can contribute to retention: test scores, performance in core subjects, whether the child lives with both parents or with a single parent, whether the parents separated during the last years of primary school, parents' labor market status, and potential job loss, and the child's illness during the school year. Controlling for these characteristics (Column

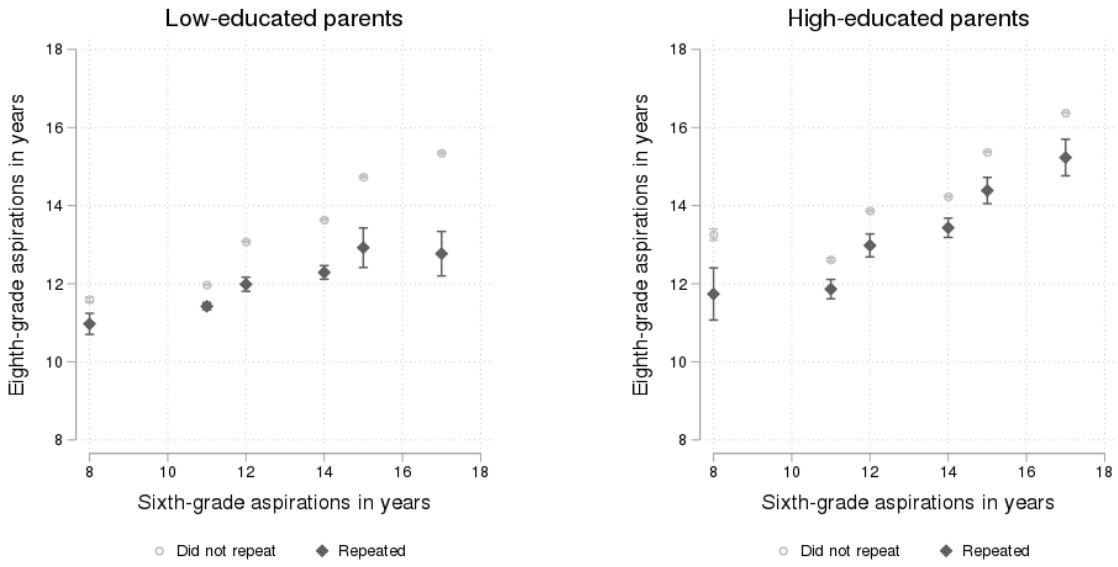

Figure 5: Change in aspired years of education by parental education and repeating

Note: Educational aspirations are presented in years. At the upper end of the aspiration distribution, a master's degree and a doctoral degree are grouped into 17 years of education. The spikes show 95 percent confidence intervals.

2), the gap between children of high-and low-educated parents decreases significantly. However, the difference still exists and is higher within schools (Column 3). Column 4, however, shows that when we look at children's sixth-grade performance in a core subject, mathematics, among those who failed at the midterm of sixth grade, children of high-educated parents are as likely to be retained as those of low-educated parents.

6.6 percent of children who failed mathematics at the midterm of sixth grade (but did not have to repeat sixth grade) had to repeat seventh grade in the end. Among those children who just passed mathematics or received at least an average evaluation, the shares of those retained are 1.9 and 0.19 percent.

Children who fail mathematics at the midterm of sixth grade most likely repeat seventh grade because they perform poorly at school. Appendix Table A2 explores the predictors of repeating within each sixth-grade midterm mathematics performance level. At the lowest level, neither marks in other subjects nor test scores increase the probability of retention; neither do parental education, labor market status, or health characteristics. The only factor that still strongly contributes to retention in this group is the sixth-grade behavior mark of the child, which measures how well the child behaves at school, according to their teachers. Whether the parents separated between the sixth and eighth grades has a large but only marginally significant effect on retention in this group.

The other two groups are more heterogeneous regarding the reason for repeating. In these

	Repeating 7th grade			
	(1)	(2)	(3)	(4)
High-educated parents	-0.00775***	-0.000848**	-0.00156***	0.0141
	(0.00040)	(0.00036)	(0.00038)	(0.0099)
6th-grade midterm mathematics performance, baseline: 1 (fail)				
2		-0.0255***	-0.0255***	-0.0191***
		(0.0048)	(0.0048)	(0.0057)
3		-0.0309***	-0.0309***	-0.0272***
		(0.0048)	(0.0048)	(0.0056)
4		-0.0311***	-0.0308***	-0.0274***
		(0.0048)	(0.0048)	(0.0056)
5		-0.0300***	-0.0293***	-0.0258***
		(0.0048)	(0.0048)	(0.0056)
High-educated parents ×				-0.0211**
2				(0.010)
High-educated parents ×				-0.0146
3				(0.0100)
High-educated parents ×				-0.0144
4				(0.0099)
High-educated parents ×				-0.0146
5				(0.0099)
Constant	0.0106***	0.100***	0.105***	0.101***
	(0.00039)	(0.023)	(0.023)	(0.023)
Observations	226903	226903	226903	226903
Controls	no	yes	yes	yes
Year fixed effect	no	yes	yes	yes
School fixed effect	no	no	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Retention probabilities by parental education and various characteristics.

Note: Linear probability models predicting retention in seventh grade with parental education and sixth-grade characteristics. The controls include variables presented in Appendix Table A1. For all variables, dummies are used for missing values. Source: Admin3 database.

groups, children of high-educated parents are ceteris paribus still less likely to get retained in seventh grade. For these groups, bad marks in another core subject, Hungarian literature, increase the risk of retention, as well as low effort and behavior grades, being raised by a single parent, parents separating at the time, and illnesses and hospitalization.

Mathematics is a subject where students' knowledge can be objectively assessed. Other subjects might give more leeway to the subjective judgment of teachers, and the advantages of higher socioeconomic background might also show in other ways: e.g., with similar lexical knowledge, children of higher-educated parents might still have better writing skills. These factors would make higher-educated parents' children less likely to retain seventh grade than those of lower-educated parents. It is also possible that after having bad marks in sixth grade but successfully passing in

the end, children of higher-educated parents catch up faster than those of lower-educated parents, so the lower chance of repeating seventh grade might already capture some compensatory advantage in catching up after a risk of retention. Therefore, if I find a compensatory advantage of high socioeconomic status in aspirations after retention, it is probably a lower bound of what I would find if children of high- and low-educated parents with similar observable characteristics had equal chances to repeat.

5.2 Compensatory advantage in eighth-grade aspirations

The following regressions explore how children of low- and high-educated parents update their educational aspirations depending on whether they were retained in seventh grade.[5] In the most detailed specification, I run the following regression:

$$\begin{aligned} \text{Asp}_{8,i,s,t} = {} & \alpha_0 + \alpha_1 \times \text{Repeated}_i + \alpha_2 \times \text{High-ed}_i + \alpha_3 \times \text{Repeated}_i \times \text{High-ed}_i \\ & + \alpha_4 \times \text{Asp}_{6,i} + \alpha_5 \times \text{Repeated}_i \times \text{Asp}_{6,i} \\ & + \alpha_6 \times \text{Asp}_{6,i} \times \text{High-ed}_i + \alpha_7 \times \text{Repeated}_i \times \text{Asp}_{6,i} \times \text{High-ed}_i \\ & + \beta_1 \times X_i + \beta_2 \times X_{6,i} + \gamma_s + \eta_t + \epsilon_{i,s,t} \end{aligned} \tag{1}$$

Sixth- and eighth-grade aspirations are measured by the years of education the child wants to achieve in these grades. REPEATED_i is 1 if child i repeated seventh grade, and 0 otherwise. HIGH-ED_i is 1 if at least one parent of the child has a high school diploma. $\text{ASP}_{6,i}$ are demeaned sixth-grade aspirations of the child. X_i are time-invariant individual characteristics, such as gender and whether the child has special education needs. $X_{6,i}$ are sixth-grade characteristics of the child, e.g., mathematics and reading test scores, marks in core subjects, and parental labor market status. γ_s are school fixed effects and η_t are year fixed effects. α_1 then shows how much repeaters of low-educated parents decrease their aspirations, and α_3 shows the compensatory advantage of high-educated parents after retention at average sixth-grade aspirations. α_5 shows whether children of low-educated parents decrease their aspirations more after retention if they had higher aspirations in sixth grade, while α_7 shows if this decrease is mitigated by having high-educated parents.

Table 4 looks at the results of Regression 1 on the whole sample, while Table 5 splits the sample by sixth-grade midterm mathematics performance. On average, non-repeaters aspire for 14.6 years – somewhat less than a college degree –, while repeaters aspire for 2.3 years less – equivalent to about a high school diploma (Column 1). Column 2 interacts retention with parental education.

[5] For these regressions, I use the National ABC database, which contains roughly twice as many observations as the Admin3, but lacks health-related variables. Appendix Tables A3-A4 present robustness checks of the same regressions on the Admin3 database, including seventh-grade GP visits and days spent in a hospital as controls.

Children of high-educated parents aspire for 1.8 more years of education than children of low-educated parents if not retained in seventh grade. Repeaters of low-educated parents aspire for 1.7 years less education than non-repeaters. The difference is slightly higher for high-educated parents' children: they aspire for two years less if retained. Since the initial aspirations of high-educated parents' children are higher, Column 3 controls for sixth-grade aspirations. Here we can see that, on average, both high- and low-educated parents' children decrease their aspirations by almost a year if they have to repeat the seventh grade.

Column 4 adds all interactions of parental education, retention and sixth-grade aspirations, while Columns 5 and 6 add individual control variables, year and school fixed effects. Column 6 shows that, at the average sixth-grade aspiration level of 14.4 years, repeaters of low-educated parents decrease their aspirations by 0.81 years by eighth grade. In contrast, this decrease is just 0.23 years for repeaters of high-educated parents. As we have seen in Figure 5, the higher the initial aspirations are of repeaters with low-educated parents, the more they decrease them by eighth grade: for each year of sixth-grade aspirations, eighth-grade aspirations are 0.16 years lower in this group. However, this effect is entirely offset for repeaters of high-educated parents: in this group, the decrease in aspirations is the same, regardless of initial aspirations.

Table 5 splits the results by sixth-grade mathematics performance. The first two columns include those who failed mathematics in the midterm of sixth grade. In this group, 7.1 percent of children of low-educated parents and 7.4 of high-educated parents repeat seventh grade. We can see in the first column that even those not retained aspire for only slightly higher than a high school diploma, 12.3 years. Those who were retained aspire for 0.4 years less. The second column includes the most detailed specification with all controls and interactions of retention, parental education and sixth-grade aspirations. In this group, repeaters with average sixth-grade aspirations do not decrease their aspirations significantly by eighth grade: the coefficient on repeating is smaller than in the whole sample (-0.2 years) and is insignificant. There is no compensatory advantage of high-educated parents at either level of initial aspirations. On the other hand, non-repeaters of high-educated parents have higher aspirations by eighth grade, and the gain is increasing in sixth-grade aspirations. In this lowest-performing group, retention in seventh grade does not seem to affect aspirations in any way. However, those who escape retention raise their aspirations by eighth grade, and low-performer children of high-educated parents benefit more from it.

The second sample, in Columns 3 and 4, includes those who just passed mathematics in the midterm of sixth grade. These children also perform poorly in mathematics, but as we saw earlier, other factors too play a role in their seventh-grade retention. Among these children, 2.6 percent of

	Eighth-grade aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-2.355***	-1.700***	-0.941***	-1.656***	-0.796***	-0.814***
	(0.043)	(0.044)	(0.042)	(0.078)	(0.072)	(0.071)
High-ed. parents		1.811***	0.815***	0.829***	0.558***	0.443***
		(0.016)	(0.0094)	(0.0095)	(0.0076)	(0.0065)
Repeated × High-ed.		-0.293***	-0.0115	0.653***	0.583***	0.577***
parents		(0.097)	(0.082)	(0.12)	(0.11)	(0.11)
Sixth-grade aspirations			0.548***	0.520***	0.338***	0.316***
			(0.0019)	(0.0028)	(0.0027)	(0.0026)
Repeated × Sixth-grade				-0.270***	-0.155***	-0.156***
aspirations				(0.023)	(0.021)	(0.021)
High-ed. parents ×				0.0527***	0.0333***	0.0203***
Sixth-grade aspirations				(0.0034)	(0.0030)	(0.0029)
Repeated × High-ed.				0.189***	0.129***	0.133***
parents × Sixth-grade aspirations				(0.042)	(0.039)	(0.038)
Constant	14.60***	13.47***	14.08***	14.05***	11.01***	11.25***
	(0.018)	(0.013)	(0.0080)	(0.0086)	(0.16)	(0.15)
Observations	369278	369278	369278	369278	369278	369278
Sixth-grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Eighth-grade aspirations

Note: The outcome variable is eighth-grade aspirations in years. Sixth-grade aspirations are demeaned. In Column 5, sixth-grade controls are gender, whether the student has special education needs, mathematics and reading test scores, mother's labor market status, father's labor market status, whether the child lived with both parents in sixth grade, a proxy for whether the parents separated between sixth and eighth grade, performance in mathematics, literature and Hungarian grammar at the midterm of sixth grade, and midterm marks in effort and behavior. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: National ABC database.

low-educated and 1.6 percent of high-educated parents' children repeat the seventh grade. High-educated parents' children in this group have a lower chance of being retained, even after controlling for various factors affecting retention. The average eighth-grade aspirations of non-retained students are slightly higher than in the lowest-performing group: 12.9 years. Repeaters are aspiring on average for about a year less of education. In Column 4, we can see that at the average initial aspiration level, repeaters aspire for half a year less education if they have low-educated parents and

	Eighth-grade aspirations in years					
	Fail		Pass		At least average	
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.372***	-0.205	-0.947***	-0.539***	-2.045***	-0.898***
	(0.093)	(0.13)	(0.053)	(0.061)	(0.091)	(0.15)
High-ed. parents		0.412***		0.465***		0.415***
		(0.067)		(0.015)		(0.0072)
Repeated × High-ed.		0.0239		0.350***		0.576***
parents		(0.25)		(0.12)		(0.20)
Sixth-grade aspirations		0.221***		0.260***		0.323***
		(0.019)		(0.0053)		(0.0031)
Repeated × Sixth-grade		-0.0289		-0.111***		-0.141***
aspirations		(0.067)		(0.028)		(0.049)
High-ed. parents ×		0.104***		0.0546***		0.0160***
Sixth-grade aspirations		(0.034)		(0.0077)		(0.0035)
Repeated × High-ed.		-0.139		0.0790		0.187**
parents × Sixth-grade aspirations		(0.12)		(0.060)		(0.080)
Constant	12.26***	10.91***	12.93***	10.38***	15.03***	11.51***
	(0.030)	(0.61)	(0.019)	(0.28)	(0.016)	(0.30)
Observations	6402	6402	61896	61896	286897	286897
Sixth-grade controls	no	yes	no	yes	no	yes
Year fixed effects	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Eighth-grade aspirations by sixth-grade mathematics performance

Note: The outcome variable is eighth-grade aspirations in years. The sample is split by sixth-grade midterm mathematics performance: those who failed mathematics (Columns 1-2), those who just passed (Columns 3-4) and those who received at least an average mark (Columns 5-6). Every odd column shows the raw difference in eighth-grade aspirations between repeaters and non repeaters, while even columns present the specifications with all controls and interactions. The control variables are the same as in Table 4. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: National ABC database.

for only 0.2 years less if they have high-educated parents. The decrease in aspirations is increasing in initial aspirations, and the difference between high- and low-educated parents' children is constant – in contrast to the whole sample findings. Children of high-educated parents seem to benefit more from escaping seventh-grade retention in this sample, too.

The last two columns include those children who received at least an average mark in mathe-

matics in the midterm of sixth grade. Retention in this group is extremely rare, with 0.5 percent among low-educated parents' children and 0.1 percent among high-educated parents' children. The reasons for repeating are also very heterogeneous in this sample. However, this sample shows the largest differences between repeaters and non-repeaters and between high- and low-educated parents' children. The raw difference between repeaters' and non-repeaters' eighth-grade aspirations is two years: non-repeaters aspire for 15 years of education, while repeaters for only 13 years. Column 6 shows that the gap in aspirations is 0.9 years for low-educated parents' children, increasing by each sixth-grade aspiration year. On the other hand, for high-educated parents' children, the gap at average sixth-grade aspirations is lower, only 0.3 years, and constant across sixth-grade aspiration levels.

Appendix Tables A3 and A4 present the results of the same regression on the smaller Admin3 sample which contains health characteristics. The regressions include the same variables as before, but I can also control for the number of visits to the general practitioner and the days spent in a hospital in seventh grade. When including health controls, the pattern seen in Figure 4 disappears: repeaters of high-educated parents decrease their aspirations less than those of low-educated parents, but the difference is constant across sixth-grade aspiration levels. This is true for all subsamples by sixth-grade midterm mathematics performance. I only find a significant compensatory advantage for children who just passed mathematics at the midterm of sixth grade. Interestingly, when controlling for health characteristics, the decrease in aspirations among repeaters also becomes insignificant in the highest-performer group. The magnitude of the coefficient for repeating is still high, though, around minus half a year, and the size of the Admin3 sample is half the size of the National ABC, so the insignificance may be due to the low number of repeaters in the sample. On the other hand, health-related variables seem to be important predictors of retention that we have to take into account.

5.3 Transition to secondary school tracks

Finally, I look at how the aspiration changes translate to changes in the probability of attending a secondary school that gives access to tertiary education. I estimate Equation 1 again, except that the outcome variable here is a dummy indicating whether the child is in a secondary school ending with a high school diploma in tenth grade or not.

Table 6 presents the results of this regression on the whole sample. As Column 1 shows, 83 percent of non-repeaters are in a secondary school in tenth grade that provides access to tertiary education, while only 35 percent of repeaters are in this type of institution. When controlling for

parental education and sixth-grade aspirations, we can see that the gap between repeaters and non-repeaters is lower among children of high-educated parents: 22 percent vs. 36 percent for low-educated parents' children. When controlling for pre-retention characteristics and school and year fixed effects, the gap at the average sixth-grade aspirations for children of low-educated parents is 19 percentage points, while for high-educated parents' children, it is zero. Repeaters of low-educated parents are also less likely to attend this institution the higher their initial aspirations were, while for high-educated parents' repeaters, this relationship is positive.

	Secondary school ending with a high school diploma					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.476***	-0.465***	-0.357***	-0.436***	-0.191***	-0.188***
	(0.017)	(0.017)	(0.017)	(0.036)	(0.034)	(0.032)
High-ed. parents		0.261***	0.136***	0.122***	0.0781***	0.0612***
		(0.0027)	(0.0022)	(0.0021)	(0.0018)	(0.0017)
Repeated × High-ed.		0.124***	0.139***	0.234***	0.192***	0.185***
parents		(0.032)	(0.029)	(0.044)	(0.041)	(0.040)
Sixth-grade aspirations			0.0738***	0.102***	0.0604***	0.0558***
			(0.00072)	(0.00081)	(0.00076)	(0.00074)
Repeated × Sixth-grade				-0.0490***	-0.0349***	-0.0349***
aspirations				(0.011)	(0.010)	(0.0099)
High-ed. parents ×				-0.0485***	-0.0404***	-0.0405***
Sixth-grade aspirations				(0.00096)	(0.00085)	(0.00084)
Repeated × High-ed.				0.100***	0.0768***	0.0764***
parents × Sixth-grade aspirations				(0.016)	(0.015)	(0.014)
Constant	0.830***	0.662***	0.736***	0.764***	0.113**	0.143***
	(0.0026)	(0.0030)	(0.0022)	(0.0021)	(0.047)	(0.046)
Observations	245996	245996	245996	245996	245996	245996
Sixth-grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Secondary school giving a high school diploma

Note: The outcome variable is a dummy indicating that the child is in a secondary school in tenth grade that ends with a high school diploma, giving access to tertiary education. Sixth-grade aspirations are demeaned. In Column 5, sixth-grade controls are gender, whether the student has special education needs, mathematics and reading test scores, mother's labor market status, father's labor market status, whether the child lived with both parents in sixth grade, a proxy for whether the parents separated between sixth and eighth grade, performance in mathematics, literature and Hungarian grammar at the midterm of sixth grade, and midterm marks in effort and behavior. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: National ABC database.

Appendix Table A5 shows that children of high-educated parents are equally less likely to attend the academic track after repeating as those of low-educated parents. This means that these children are more likely to attend a technical secondary school, which still provides access to tertiary education. In contrast, repeaters of low-educated parents are more likely to attend the vocational track, which does not lead to tertiary education. Appendix Table A6 also shows that repeaters are more likely than non-repeaters to be missing from the tenth-grade sample – this can be because they had to repeat multiple grades subsequently and did not reach the tenth grade by the end of the sample period. However, it can also mean they dropped out of school before the tenth grade. Missing from the tenth-grade sample is also more prevalent for repeaters of low-educated parents, and the gap by parental education is increasing in initial aspirations.

Table 7 splits the sample by sixth-grade mathematics performance. Although the aspirations of the worst-performer repeaters were not lower, they are less likely to attend a secondary school track that leads to tertiary education than non-repeaters. The gap is increasing in sixth-grade aspirations but not significantly lower for children of high-educated parents. However, even in this low-performer group, among non-repeaters, high-educated parents' children are 13 percentage points more likely to attend a secondary school ending with a high school diploma. In the group that just passed mathematics in the midterm of sixth grade, repeaters are only marginally significantly less likely to attend this type of institution, and there is no significant difference between low- and high-educated parents' children. On the other hand, non-repeaters in this group are still 12 percentage points more likely to attend this track if they have high-educated parents. The patterns in the best-performing group are similar to those in aspirations: Repeaters of low-educated parents with average sixth-grade aspirations are 23 percentage points less likely to attend a secondary school ending with a high school diploma than non-repeaters. High-educated parents almost entirely offset this difference: the gap is only four percentage points for their children. The gap between non-repeaters and repeaters is not increasing in sixth-grade aspirations for children of low-educated parents – in contrast to eighth-grade aspirations – but is decreasing for children of high-educated parents.

Appendix Tables A7 and A8 present the robustness checks that include health-related controls as well. The patterns in the whole sample are somewhat different: repeaters of low-educated parents with average sixth-grade aspirations are 12 percentage points less likely to attend a secondary school ending with a high school diploma than non-repeaters. However, high-educated parents do not significantly offset this gap. Although the gap is not increasing in sixth-grade aspirations for repeaters of low-educated parents, it is decreasing significantly for those of high-educated parents.

	Secondary school ending with a high school diploma					
	Fail		Pass		At least average	
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.154***	-0.106**	-0.204***	-0.0603*	-0.329***	-0.233***
	(0.030)	(0.050)	(0.023)	(0.033)	(0.034)	(0.068)
High-ed. parents		0.129***		0.120***		0.0409***
		(0.030)		(0.0056)		(0.0016)
Repeated × High-ed.		0.0516		0.0641		0.193**
parents		(0.11)		(0.050)		(0.077)
Sixth-grade aspirations		0.0349***		0.0521***		0.0487***
		(0.0087)		(0.0019)		(0.00088)
Repeated × Sixth-grade		-0.0573***		-0.0149		-0.0302
aspirations		(0.022)		(0.016)		(0.022)
High-ed. parents ×		0.0206		0.000192		-0.0390***
Sixth-grade aspirations		(0.014)		(0.0027)		(0.00097)
Repeated × High-ed.		0.0527		0.0242		0.0806***
parents × Sixth-grade aspirations		(0.055)		(0.026)		(0.029)
Constant	0.317***	0.0398	0.511***	-0.303***	0.903***	0.406***
	(0.0097)	(0.25)	(0.0052)	(0.095)	(0.0018)	(0.089)
Observations	3350	3350	37213	37213	196942	196942
Sixth-grade controls	no	yes	no	yes	no	yes
Year fixed effects	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Secondary school giving a high school diploma by sixth-grade mathematics performance

Note: The outcome variable is a dummy indicating that the child is in a secondary school in tenth grade that ends with a high school diploma, giving access to tertiary education. The sample is split by sixth-grade midterm mathematics performance: those who failed mathematics (Columns 1-2), those who just passed (Columns 3-4) and those who received at least an average mark (Columns 5-6). Every odd column shows the raw difference in eighth-grade aspirations between repeaters and non repeaters, while even columns present the specifications with all controls and interactions. The control variables are the same as in Table 4. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: National ABC database.

The patterns for the subsamples in Appendix Table A8 are similar to the ones in Table 7. However, most coefficients are not significant, probably due to the even smaller sample size than in the aspiration regressions.[6]

[6]I could not link tenth-grade data to around one-third of the sample, so the sample sizes in the secondary school track regressions are only two-thirds of those of the aspiration regressions.

6 Discussion

I found that children of low-educated parents start with lower educational aspirations than children of high-educated parents with similar skills. Although grade retention by the end of primary school is already rare in Hungary, it is much more frequent among children of low-educated parents, who already have lower aspirations and worse secondary school track prospects. Grade retention is detrimental to aspirations and later secondary school track choice, and on average, it affects children of low-educated parents more adversely, increasing the – already high – aspiration and secondary education gaps.

The relationship between retention and aspirations is heterogenous across the presumable reasons of repeating. Those children who perform poorly in mathematics – a core and relatively objectively evaluable subject –, do not lose from retention in terms of aspirations but are less likely to end up in a secondary school that gives access to tertiary education than those who did not have to repeat seventh grade in the end. These differences are not significantly different across parental backgrounds. On the other hand, those who escaped retention increase their aspirations and chances of attending a secondary track leading to tertiary education more if they have high-educated parents.

For those who perform better in mathematics, various factors affect retention additionally: marks in other subjects, whether they live with both parents and whether the parents separated around that time or health characteristics. In these groups, the compensatory advantage of high-educated parents is more prevalent after retention: it seems to be there both in aspirations and subsequent secondary school tracks. Parents try to avoid downward mobility by pushing children towards at least as high education levels as their own. After a negative shock, such as retention, they may still try to push their children toward better education, but higher-educated parents have more resources to do so, and they also start from a higher reference point. If this is the case, it might be worth working on policies that help children of low-educated parents catch up after having to repeat a grade or, even better, help them avoid retention.

References

Agasisti, Tommaso and Konstantina Maragkou, "Socio-economic gaps in educational aspirations: do experiences and attitudes matter?," *Education Economics*, 2023, *31* (4), 471–487.

Akerlof, George A. and Rachel E. Kranton, "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature*, December 2002, *40* (4), 1167–1201.

Andrew, M., "The Scarring Effects of Primary-Grade Retention? A Study of Cumulative Advantage in the Educational Career," *Social Forces*, 2014, *93* (2), 653–685.

Árendás, Zsuzsanna, Judit Durst, Noémi Katona, and Vera Messing, "The Limits of Trading Cultural Capital: Returning Migrant Children and their Educational Trajectory in Hungary," in Adrienne Lee Atterberry, Derrace Garfield McCallum, Siqi Tu, and Amy Lutz, eds., *Children and Youths' Migration in a Global Landscape*, Emerald Publishing Limited, 2022, pp. 115–139.

Bernardi, Fabrizio, "Compensatory Advantage as a Mechanism of Educational Inequality: A Regression Discontinuity Based on Month of Birth," *Sociology of Education*, 2014, *87* (2), 74–88.

_ **and Héctor Cebolla-Boado**, "Previous School Results and Social Background: Compensation and Imperfect Information in Educational Transitions," *European Sociological Review*, 10 2013, *30* (2), 207–217.

_ **and Manuel T. Valdés**, "Sticky educational expectations: A cross-country comparison," *Research in Social Stratification and Mobility*, 2021, *75*, 100624.

_ **and Michael Grätz**, "Making up for an unlucky month of birth in school : causal evidence on the compensatory advantage of family background in England," *Sociological Science*, 2015, *2*, 235–251.

_ **and Moris Triventi**, "Compensatory advantage in educational transitions: Trivial or substantial? A simulated scenario analysis," *Acta Sociologica*, 07 2018, *63*, 000169931878095.

Cham, Heining, Jan N. Hughes, Stephen G. West, and Myung Hee Im, "Effect of retention in elementary grades on grade 9 motivation for educational attainment," *Journal of School Psychology*, 2015, *53* (1), 7–24.

Contini, Dalit and Guido Salza, "Children left behind. New evidence on the (adverse) impact of grade retention on educational careers.," Working Paper 2022.

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers, "Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences," *Journal of Political Economy*, 2019.

Ganzeboom, Harry, Paul Graaf, and Donald Treiman, "A Standard Socio-Economic Index of Occupational Status," *Social Science Research*, 03 1992, *21*, 1–56.

Genicot, Garance and Debraj Ray, "Aspirations and Inequality," *Econometrica*, 2017, *85* (2), 489–519.

_ and _, "Aspirations and economic behavior," *Annual Review of Economics*, 2020, *12*, 715–746.

Guyon, Nina and Elise Huillery, "Biased Aspirations and Social Inequality at School: Evidence from French Teenagers," *The Economic Journal*, 06 2020, *131* (634), 745–796.

Heath, Chip, Richard P. Larrick, and George Wu, "Goals as Reference Points," *Cognitive Psychology*, 1999, *38*, 79–109.

Heckman, James J. and Stefano Mosso, "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 2014, *6* (1), 689–733.

Horn, Dániel, "Diverging performances: the detrimental effects of early educational selection on equality of opportunity in Hungary," *Research in Social Stratification and Mobility*, 2013, *32*, 25–43. Social mobility and inequality in the life course: Exploring the relevance of context.

Hughes, Jan N., Oi-Man Kowk, and Im MyungHee, "Effect of Retention in First Grade on Parents' Educational Expectations and Children's Academic Outcomes," *American Educational Research Journal*, 2013, *50* (6).

OECD, *Education at a Glance 2015* 2015.

Page, Lionel, Louis Levy Garboua, and Claude Montmarquette, "Aspiration levels and educational choices: An experimental study," *Economics of Education Review*, 2007, *26* (6), 747–757.

Sebők, Anna, "The Panel of Linked Administrative Data of CERS Databank," *Budapest Working Papers on the Labour Market*, 2019, (2).

A Appendix

A.1 Parental background by mathematics test scores

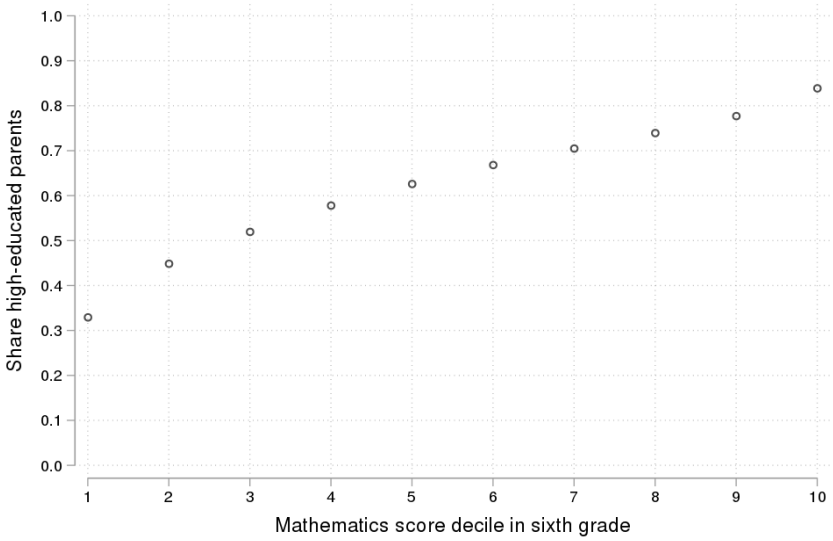

Figure A1: Share of children with high-educated parents by mathematics test scores

Note: The sample includes children who spent all eight years in primary school and were not retained until the seventh grade. High-educated parents mean at least one parent with a high school diploma.

A.2 Tenth-grade secondary school track by eighth-grade aspirations

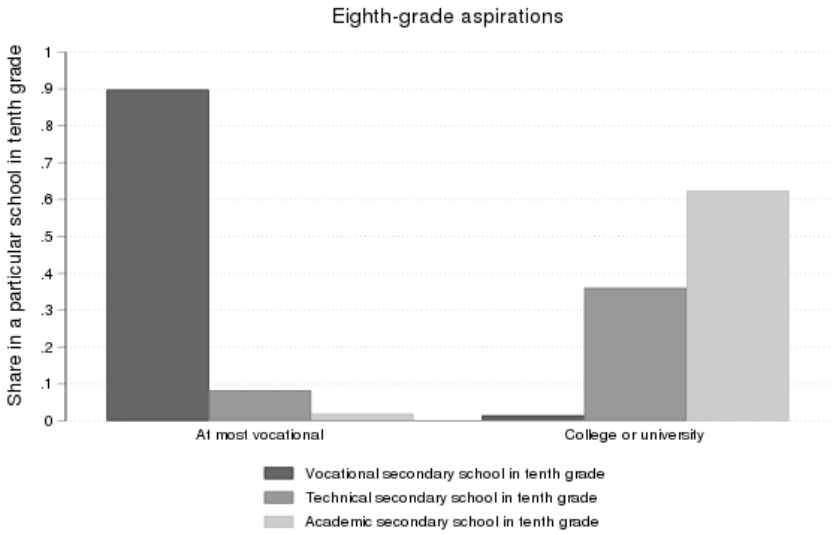

Figure A2: Share of students in secondary school tracks by eighth-grade aspirations

Note: The sample includes children who spent all eight years in primary school and were not retained until the seventh grade. Note that by the time of stating their aspirations in eighth grade, most children already know the secondary school to which they were admitted.

A.3 Missing data

As completing the student background questionnaire is voluntary, there is a selection bias in the sample if completion is nonrandom. In the analysis, I include observations with missing values by using dummies. However, I cannot apply this technique to parental education and the outcome variables. Six percent of the sample has missing parental education data, so I exclude these observations from the analyses. The average sixth-grade mathematics test score in this sample is 1447 points, compared to 1407 for children of low-educated parents and 1546 for those of high-educated parents. The share of repeaters among children with missing parental education data is 5.2 percent, while it is 1.9 in the rest of the sample. Only 20 percent of this sample with missing data has data on eighth-grade educational aspirations. On average, these children aspire for 13.7 years of education, compared to 13.3 years for children of low-educated parents and 15.4 for children of high-educated parents. These comparisons suggest a selection bias in completion towards children who perform better at school, have higher aspirations, and probably have higher-educated parents. Therefore, the results might underestimate the compensatory advantage of high socioeconomic background, as I lose many observations of children from lower socioeconomic backgrounds.

A.4 Characteristics of repeaters and non-repeaters

Children of different parental backgrounds have unequal chances of repeating the seventh grade at similar sixth-grade school performance. Therefore, I explore who the repeaters are in each group and what are the predictors of their retention. To compare repeaters and non-repeaters by a broader range of characteristics, I use the administrative Admin3 database. Table A1 summarizes the characteristics of repeaters and non-repeaters by parental education. Since the size of the administrative database is only half of the National ABC, the numbers of repeaters in both groups are also relatively low. I compare factors that might affect retention: sixth-grade test scores and midterm marks in different subjects, whether the parents lived together in sixth grade, and whether they potentially separated between sixth and eighth grades (measured by the child living with both parents in sixth grade but living with only one of them in eighth grade), parental labor market status and its changes between sixth and eighth grades, visits to the general practitioner, and days spent in hospital. In general, repeaters of both low- and high-educated parents perform worse in sixth grade than non-repeaters. However, the average performance of children with high-educated parents is better in all subjects and standardized tests. There are some differences regarding the family structure: repeaters with high-educated parents are more likely to live with a single parent than repeaters with low-educated parents, suggesting that being raised by a single parent may put a higher risk of retention on them. The share of separated parents is higher among repeaters in both groups. Another factor that might be a more frequent cause for retention for children of high-educated parents is hospitalization: repeaters in this group spend on average twice as much in a hospital as repeaters of low-educated parents (1.54 vs. 0.75 days), while the difference is relatively small for non-repeaters (0.24 vs. 0.28 days).

	Low - Did not repeat	Low - Repeated	High - Did not repeat	High - Repeated
Sixth-grade maths test score	1439.52	1318.80	1551.94	1398.29
	(172.39)	(146.14)	(170.33)	(162.66)
Sixth-grade reading test score	1427.96	1296.19	1551.00	1373.66
	(173.66)	(148.57)	(171.97)	(174.12)
Sixth-grade midterm maths mark	3.17	2.06	3.84	2.32
	(1.01)	(0.73)	(0.96)	(0.94)
Sixth-grade midterm literature	3.61	2.48	4.23	2.83
mark	(0.98)	(0.81)	(0.83)	(1.01)
Sixth-grade midterm grammar	3.41	2.39	4.01	2.80
mark	(0.96)	(0.77)	(0.88)	(0.98)
Lives with both parents in	0.77	0.67	0.78	0.49
sixth grade	(0.42)	(0.47)	(0.41)	(0.50)
Parents separated between	0.05	0.11	0.04	0.13
grades six and eight	(0.22)	(0.31)	(0.21)	(0.33)
Mother does not work in sixth	0.37	0.48	0.15	0.23
grade	(0.48)	(0.50)	(0.36)	(0.42)
Mother works full-time in	0.49	0.40	0.74	0.63
sixth grade	(0.50)	(0.49)	(0.44)	(0.48)
Father does not work in sixth	0.15	0.24	0.07	0.13
grade	(0.36)	(0.43)	(0.25)	(0.33)
Father works full-time in	0.64	0.54	0.72	0.65
sixth grade	(0.48)	(0.50)	(0.45)	(0.48)
Mother stopped working between	0.07	0.09	0.04	0.09
grades six and eight	(0.26)	(0.29)	(0.20)	(0.29)
Father stopped working between	0.06	0.08	0.03	0.04
grades six and eight	(0.23)	(0.26)	(0.17)	(0.20)
Visits to GP in seventh grade	4.99	7.58	4.09	6.33
	(4.82)	(6.74)	(4.10)	(6.91)
Days spent in a hospital in	0.28	0.76	0.25	1.57
seventh grade	(1.76)	(4.82)	(1.72)	(9.10)
Observations	53689	492	100999	212

Table A1: Summary statistics of repeaters and non-repeaters by parental education

Note: The first two columns include children with low-educated parents, while the last two include those with high-educated parents. Rows 6-13 show shares of students with a particular characteristic; the rest of the rows show average levels. Marks in different subjects are given from 1 to 5, with 1 meaning 'fail' and 5 'excellent'. A parent not working means he/she is on childcare allowance, unemployed, retired, permanently ill/disabled, or does not work for another reason. A parent having stopped working means he/she moved from working in any type of job to not working because of either of the reasons above. Source: Admin3 database.

	Repeated 7th grade		
	(1)	(2)	(3)
	Fail	Pass	Higher marks
High-educated parents	0.00361	-0.00861***	-0.000542*
	(0.018)	(0.0018)	(0.00030)
Sixth-grade maths test score	0.0000211	-0.0000213***	-0.00000310***
	(0.000056)	(0.0000071)	(0.00000096)
Sixth-grade reading test score	-0.0000958	-0.0000237***	-0.00000102
	(0.000068)	(0.0000068)	(0.0000011)
Sixth-grade midterm literature	-0.0158	-0.00489***	-0.00121***
mark	(0.012)	(0.0013)	(0.00027)
Sixth-grade midterm grammar	0.00163	-0.00214*	0.000310
mark	(0.011)	(0.0013)	(0.00022)
6th year midterm effort grade	-0.00496	-0.00922***	-0.00101***
	(0.012)	(0.0015)	(0.00031)
6th year midterm behavior	-0.0294***	-0.00948***	-0.00163***
grade	(0.0094)	(0.0012)	(0.00027)
Lives with both parents in	-0.0205	-0.00799***	-0.000885***
sixth grade	(0.016)	(0.0019)	(0.00031)
Parents separated between	0.0573*	0.0171***	0.00149**
grades six and eight	(0.035)	(0.0045)	(0.00071)
Mother does not work in sixth	-0.00347	-0.000233	0.000613
grade	(0.022)	(0.0027)	(0.00047)
Mother works full-time in	-0.0127	0.000252	-0.000169
sixth grade	(0.024)	(0.0027)	(0.00036)
Father does not work in sixth	0.0368	0.00386	0.000635
grade	(0.024)	(0.0032)	(0.00056)
Father works full-time in	0.00980	-0.000499	-0.000289
sixth grade	(0.020)	(0.0021)	(0.00028)
Mother stopped working between	-0.00690	0.00602	0.00121*
grades six and eight	(0.030)	(0.0038)	(0.00073)
Father stopped working between	-0.0156	-0.00622*	0.000336
grades six and eight	(0.034)	(0.0035)	(0.00077)
Visits to GP in seventh grade	0.00245	0.00161***	0.000186***
	(0.0016)	(0.00025)	(0.000050)
Days spent in a hospital in	-0.00272	0.00227**	0.00110***
seventh grade	(0.0025)	(0.0011)	(0.00041)
Constant	0.294***	0.164***	0.0237***
	(0.10)	(0.013)	(0.0020)
Observations	3286	32999	165823
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Probability of retention by sixth-grade midterm mathematics performance.

Note: Source: Admin3 database. The columns report the results of linear probability models predicting retention in seventh grade in three levels of sixth-grade mathematics performance: fail (1), pass (2), and higher marks (3-5).

A.5 Eighth-grade aspiration regressions with health controls

	Eighth-grade aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-2.696***	-1.896***	-1.024***	-1.692***	-0.744***	-0.753***
	(0.077)	(0.071)	(0.068)	(0.16)	(0.15)	(0.15)
High-ed. parents		2.148***	1.029***	1.062***	0.636***	0.495***
		(0.023)	(0.015)	(0.016)	(0.014)	(0.013)
Repeated × High-ed.		-0.282	0.000546	0.609**	0.634***	0.573**
parents		(0.18)	(0.16)	(0.25)	(0.24)	(0.24)
Sixth-grade aspirations			0.531***	0.493***	0.327***	0.311***
			(0.0028)	(0.0048)	(0.0047)	(0.0047)
Repeated × Sixth-grade				-0.204***	-0.119***	-0.116***
aspirations				(0.044)	(0.042)	(0.042)
High-ed. parents ×				0.0588***	0.0750***	0.0681***
Sixth-grade aspirations				(0.0056)	(0.0052)	(0.0052)
Repeated × High-ed.				0.128	0.0459	0.0368
parents × Sixth-grade aspirations				(0.087)	(0.082)	(0.081)
Constant	15.12***	13.75***	14.46***	14.41***	10.38***	10.62***
	(0.023)	(0.017)	(0.012)	(0.015)	(0.27)	(0.27)
Observations	181426	181426	181426	181426	181426	181426
Sixth-grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Eighth-grade aspirations

Note: The outcome variable is eighth-grade aspirations in years. Sixth-grade aspirations are demeaned. In Column 5, sixth-grade controls are gender, whether the student has special education needs, mathematics and reading test scores, mother's labor market status, father's labor market status, whether the child lived with both parents in sixth grade, a proxy for whether the parents separated between sixth and eighth grade, performance in mathematics, literature and Hungarian grammar at the midterm of sixth grade, midterm marks in effort and behavior, number of visits to the GP in seventh grade, and days spent in hospital in seventh grade. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: Admin3 database.

	Eighth-grade aspirations in years					
	Fail		Pass		At least average	
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.418**	-0.234	-0.995***	-0.537***	-2.244***	-0.534
	(0.18)	(0.33)	(0.083)	(0.089)	(0.19)	(0.35)
High-ed. parents		0.496***		0.500***		0.476***
		(0.16)		(0.027)		(0.016)
Repeated × High-ed.		-0.0952		0.540**		0.133
parents		(0.65)		(0.21)		(0.46)
Sixth-grade aspirations		0.239***		0.233***		0.326***
		(0.053)		(0.010)		(0.0055)
Repeated × Sixth-grade		-0.141		-0.0803**		-0.0521
aspirations		(0.17)		(0.036)		(0.090)
High-ed. parents ×		0.137		0.0616***		0.0629***
Sixth-grade aspirations		(0.093)		(0.016)		(0.0061)
Repeated × High-ed.		-0.0491		0.0399		0.0377
parents × Sixth-grade aspirations		(0.39)		(0.13)		(0.14)
Constant	12.43***	10.09***	13.07***	9.718***	15.58***	11.10***
	(0.046)	(0.86)	(0.022)	(0.45)	(0.022)	(0.46)
Observations	2760	2760	27966	27966	144481	144481
Sixth-grade controls	no	yes	no	yes	no	yes
Year fixed effects	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Eighth-grade aspirations by sixth-grade mathematics performance

Note: The outcome variable is eighth-grade aspirations in years. The sample is split by sixth-grade midterm mathematics performance: those who failed mathematics (Columns 1-2), those who just passed (Columns 3-4) and those who received at least an average mark (Columns 5-6). Every odd column shows the raw difference in eighth-grade aspirations between repeaters and non repeaters, while even columns present the specifications with all controls and interactions. The control variables are the same as in Table A3. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: Admin3 database.

A.6 Secondary school tracks

	Academic secondary school					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.308***	-0.176***	-0.0357***	-0.230***	-0.0775***	-0.0643***
	(0.010)	(0.0083)	(0.010)	(0.017)	(0.018)	(0.017)
High-ed. parents		0.304***	0.141***	0.155***	0.111***	0.0843***
		(0.0041)	(0.0029)	(0.0030)	(0.0028)	(0.0022)
Repeated × High-ed.		-0.149***	-0.129***	0.0503	0.0553*	0.0282
parents		(0.023)	(0.023)	(0.031)	(0.030)	(0.030)
Sixth-grade aspirations			0.0956***	0.0661***	0.0273***	0.0232***
			(0.00078)	(0.00084)	(0.00078)	(0.00074)
Repeated × Sixth-grade				-0.0612***	-0.0338***	-0.0321***
aspirations				(0.0053)	(0.0058)	(0.0054)
High-ed. parents ×				0.0511***	0.0389***	0.0308***
Sixth-grade aspirations				(0.0011)	(0.0010)	(0.00098)
Repeated × High-ed.				0.00858	0.0111	0.00529
parents × Sixth-grade aspirations				(0.012)	(0.012)	(0.011)
Constant	0.404***	0.209***	0.305***	0.275***	-0.275***	-0.231***
	(0.0043)	(0.0028)	(0.0028)	(0.0032)	(0.047)	(0.046)
Observations	245996	245996	245996	245996	245996	245996
Sixth-grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Being in an academic secondary school in tenth grade

Note: The outcome variable is a dummy indicating that the child is in an academic secondary school in tenth grade. Sixth-grade aspirations are demeaned. Control variables are the same as in Table 6. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: National ABC database.

	Not found in the tenth-grade sample					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	0.284***	0.306***	0.293***	0.317***	0.330***	0.333***
	(0.011)	(0.013)	(0.013)	(0.020)	(0.021)	(0.021)
High-ed. parents		-0.0507***	-0.0336***	-0.0286***	-0.0187***	-0.0154***
		(0.0027)	(0.0024)	(0.0024)	(0.00092)	(0.00092)
Repeated × High-ed.		-0.121***	-0.126***	-0.165***	-0.107***	-0.111***
parents		(0.024)	(0.024)	(0.031)	(0.028)	(0.028)
Sixth-grade aspirations			-0.00939***	-0.0184***	-0.0102***	-0.00876***
			(0.00054)	(0.00076)	(0.00051)	(0.00047)
Repeated × Sixth-grade				0.0148**	0.00356	0.00332
aspirations				(0.0061)	(0.0064)	(0.0064)
High-ed. parents ×				0.0162***	0.0114***	0.0102***
Sixth-grade aspirations				(0.00094)	(0.00054)	(0.00050)
Repeated × High-ed.				-0.0386***	-0.0217**	-0.0213**
parents × Sixth-grade aspirations				(0.011)	(0.0097)	(0.0097)
Constant	0.332***	0.364***	0.353***	0.343***	0.212***	0.194***
	(0.0023)	(0.0029)	(0.0027)	(0.0027)	(0.032)	(0.032)
Observations	369278	369278	369278	369278	369278	369278
Sixth-grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Not being found in the tenth-grade sample

Note: The outcome variable is a dummy that is 1 if I did not find the child in the tenth-grade sample. Sixth-grade aspirations are demeaned. Control variables are the same as in Table 6. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: National ABC database.

A.6.1 With health controls

	Secondary school ending with a high school diploma					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.503***	-0.488***	-0.422***	-0.378***	-0.124**	-0.121***
	(0.026)	(0.027)	(0.026)	(0.048)	(0.048)	(0.046)
High-ed. parents		0.237***	0.162***	0.141***	0.0784***	0.0615***
		(0.0030)	(0.0027)	(0.0026)	(0.0023)	(0.0023)
Repeated × High-ed.		0.109**	0.116**	0.104*	0.101*	0.0948
parents		(0.050)	(0.046)	(0.060)	(0.059)	(0.058)
Sixth-grade aspirations			0.0378***	0.0650***	0.0350***	0.0324***
			(0.00056)	(0.00085)	(0.00074)	(0.00073)
Repeated × Sixth-grade				-0.000970	0.00142	0.00153
aspirations				(0.012)	(0.012)	(0.011)
High-ed. parents ×				-0.0406***	-0.0295***	-0.0290***
Sixth-grade aspirations				(0.00091)	(0.00079)	(0.00078)
Repeated × High-ed.				0.0543***	0.0361**	0.0356**
parents × Sixth-grade aspirations				(0.016)	(0.016)	(0.016)
Constant	0.852***	0.697***	0.741***	0.773***	0.0369	0.0480
	(0.0024)	(0.0032)	(0.0027)	(0.0025)	(0.055)	(0.056)
Observations	127441	127441	127441	127441	127441	127441
Sixth-grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Secondary school giving a high school diploma

Note: The outcome variable is a dummy indicating that the child is in a secondary school in tenth grade that ends with a high school diploma, giving access to tertiary education. Sixth-grade aspirations are demeaned. In Column 5, sixth-grade controls are gender, whether the student has special education needs, mathematics and reading test scores, mother's labor market status, father's labor market status, whether the child lived with both parents in sixth grade, a proxy for whether the parents separated between sixth and eighth grade, performance in mathematics, literature and Hungarian grammar at the midterm of sixth grade, midterm marks in effort and behavior, number of visits to the GP in seventh grade, and days spent in hospital in seventh grade. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: Admin3 database.

	Secondary school ending with a high school diploma					
	Fail		Pass		At least average	
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-0.112**	-0.173	-0.269***	-0.0644	-0.354***	-0.108
	(0.052)	(0.12)	(0.035)	(0.051)	(0.050)	(0.074)
High-ed. parents		0.185***		0.120***		0.0458***
		(0.069)		(0.0088)		(0.0022)
Repeated × High-ed.		-0.230		-0.0213		0.0136
parents		(0.32)		(0.076)		(0.094)
Sixth-grade aspirations		0.0482**		0.0411***		0.0258***
		(0.021)		(0.0029)		(0.00076)
Repeated × Sixth-grade		-0.130**		-0.00660		0.0272*
aspirations		(0.056)		(0.026)		(0.014)
High-ed. parents ×		-0.0204		-0.00633*		-0.0239***
Sixth-grade aspirations		(0.031)		(0.0037)		(0.00081)
Repeated × High-ed.		0.230		0.0148		0.00177
parents × Sixth-grade aspirations		(0.18)		(0.030)		(0.023)
Constant	0.343***	0.00843	0.536***	-0.733***	0.914***	0.501***
	(0.015)	(0.37)	(0.0058)	(0.12)	(0.0017)	(0.098)
Observations	1425	1425	17380	17380	104694	104694
Sixth-grade controls	no	yes	no	yes	no	yes
Year fixed effects	no	yes	no	yes	no	yes
School fixed effects	no	yes	no	yes	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Secondary school giving a high school diploma by sixth-grade mathematics performance

Note: The outcome variable is a dummy indicating that the child is in a secondary school in tenth grade that ends with a high school diploma, giving access to tertiary education. The sample is split by sixth-grade midterm mathematics performance: those who failed mathematics (Columns 1-2), those who just passed (Columns 3-4) and those who received at least an average mark (Columns 5-6). Every odd column shows the raw difference in eighth-grade aspirations between repeaters and non repeaters, while even columns present the specifications with all controls and interactions. The control variables are the same as in Table A7. For all variables, dummies are used for missing values. Standard errors are clustered on the sixth-grade school level. Source: Admin3 database.

A.7 Data Appendix

The student-level questionnaires and test score data are available between 2008 and 2017, totaling around 2.8 million in observations. I kept students for whom I could link at least sixth- and eighth-grade data. Then, I created an indicator variable for whether the student repeated any time up to the sixth grade. I used two variables here: whether the student repeated in lower grades of primary school and whether she repeated in the higher grades, both reported in grade six. I classified the student as non-repeater if both variables equal one, which means no repetition in those grades. Unfortunately, 20 percent of this constructed measure was missing in sixth grade. I recovered this data from the eighth-grade and tenth-grade surveys, wherever possible since later surveys also include these questions. When reporting about repetition in lower grades was inconsistent through survey years, I used the mode of the answers to impute the sixth-grade value. In the end, I had self-reported information on repetition up to the sixth grade from 94.3 percent of the sixth-grade sample. Because of Hungary's compulsory school starting age rules, students are 12 or 13 years old at the end of the sixth grade (when the test takes place). However, in a small percentage of cases, it is possible that students – without repeating a grade – turn 14 in that year. Therefore, I dropped students younger than 12 when writing the sixth-grade test and assigned students older than 14 to the repeater group.[7] This is around 1 percent of the sample. I constructed a wide database where the time variable is grade, and I kept the first observations from the sixth, eighth and tenth grades.

A.7.1 Creating years of education from the categorical education variables

I created a continuous variable from educational aspirations with the following coding: unfinished primary education = seven years, primary education = eight years, vocational and technical secondary education = 11 years, high school diploma = 12 years, upper-secondary non-tertiary qualification = 14 years, college or BA degree = 15 years, university or MA degree and doctoral degree = 17 years. I chose to code a doctoral degree similar to a master's degree because, in primary school, children probably cannot apprehend the length of a doctoral education, so setting their aspirations to 20-22 years of schooling would be an overestimation of their educational preferences.

A.7.2 Admin3 NABC extension

I took the same data-cleaning steps as with the National ABC. I first matched the observations from the National ABC extension to data from May of the same year in the administrative database. I

[7] I used the first occurrence in the sixth-grade sample since a fraction of students had to repeat sixth grade, which resulted in them appearing multiple times in the sixth-grade samples.

dropped about 280 observations where the matches were wrong: these ID-s belonged to much older people in the Admin than in the National ABC. I dropped everyone older than 22 (the highest age in the National ABC) in the matched database. I checked if the rest of the observations were matched well in terms of age: I considered a match bad if the age from the Admin3 database were higher than age + 1 or lower than age - 1 from the National ABC database. First, I corrected the birth year for those observations where the age mismatch resulted from inconsistent reporting of the birth year. In the end, I still had 40 observations with age mismatch, so I dropped these observations. I did the same steps as with the National ABC database, but some data were missing here: there were no data on class and special education needs and valid/ not valid test status. I merged the National ABC data to the monthly Admin3 database by setting the NABC month as May. Then I merged the monthly healthcare data to the main file. For the estimations, I created school-year aggregates of the monthly healthcare data, e.g., the number of visits to the GP in the sixth grade, seventh grade, etc. I kept the first observations from the sixth-, eighth-, and tenth-grade NABC data and the corresponding data from Admin3 from these years. I also kept the seventh-grade healthcare data for predicting retention.