Compensatory advantage and inequality in educational aspirations*

Flóra Drucker†

2022 September

Preliminary draft

Click here for the latest version.

Abstract

There is a large body of evidence supporting that family background determines the development of children from early on. Children from higher and lower socioeconomic backgrounds enter the education system with different cognitive and non-cognitive skills. Besides differences in early childhood, a child's socioeconomic status (SES) can also affect how she copes with hardships at later stages in her educational career. Using two large, administrative and survey-based, datasets from Hungary, I look at how children from different socioeconomic backgrounds change their educational aspirations after one specific hardship – grade retention in the 7^{th} grade of primary school. Using the difference-in-differences method I find that children from all socioeconomic backgrounds decrease their aspirations after retention, but the magnitudes are larger for low-SES children. The post-retention SES gap in aspirations is the highest for those children who had high aspirations before retention. However, when looking at subsamples of children by 6^{th}-grade mathematics performance, the effects are heterogeneous. For those children who failed in mathematics at the mid-term of 6^{th} grade – so those who most likely repeated 7^{th} grade due to their low performance at school –, there is no compensatory advantage of high socioeconomic background either in 8^{th}-grade aspirations or in transitioning to a secondary school that gives access to tertiary education. For children who had better grades in mathematics in their 6^{th} year, there is, on the other hand, a compensatory advantage in both outcomes.

*I thank Attila Gáspár, Dániel Horn, Botond Kőszegi, Alexander Cappelen, Mats Köster and Tamás Keller for discussions and helpful suggestions. I also thank participants of various conferences and seminars for their comments. I am grateful for the Databank of the Centre for Economic and Regional Studies for providing access to the databases.

†Central European University

1 Introduction

There is high correlation between the education levels of parents and their children. In 2012, 52% of 25-32 year-olds had the same education level as their parents on average in the OECD countries (OECD, 2015). Parental education influences children's education outcomes through several channels. First, parental background affects cognitive and non-cognitive skills of children already in early ages. Parental investment in early childhood is crucial (see a review in Heckman and Mosso, 2014). Falk et al. (2019) find that children from high socioeconomic background are more intelligent, more altruistic, less risk-seeking, and more patient already at ages 7-9. The main channels through which parental social status affects children's IQ and preferences in early childhood are time spent with the children and parenting style. Whether these differences remain or diminish depends largely on the education system. School systems which track students into academic or vocational tracks later can increase intergenerational educational mobility by decreasing inequality in the school performance of students (see e.g., Pekkarinen, 2018; Schütz et al., 2008). Besides entering school with different skills, children from low and high socioeconomic status (SES) also differ in their resources to cope with hardships during their educational career. A strand of literature in sociology explores a specific channel of low intergenerational educational mobility: the compensatory advantage of high socioeconomic status. After a negative school event – like failing a subject or a bad school choice –, high-SES students have much more resources to compensate for this negative shock, and they recover from the shock faster than low-SES students. This compensatory advantage reinforces the initial differences in socioeconomic status (Bernardi, 2014; Bernardi and Cebolla-Boado, 2013; Bernardi and Grätz, 2015).

I am addressing 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 and 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. Grade retention is originally there to give a second chance to students who failed one or more subjects for the first time. Therefore, it could be a way to decrease the inequality in the performance of students, 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. The existing evidence on the effects of grade retention in later years of primary school is inconclusive. Jacob and Lefgren (2004) and Jacob and Lefgren (2009) use a regression discontinuity

design and a standardized test-based promotion system in the US to look at short-run and long-run effects of grade retention. They find that retention does not have a consistent effect in the short run, and retaining students in the 6^{th} grade of primary school does not have long-run effects – on high school completion–, either. However, retention in 8^{th} grade increases the probability of dropping out of high school. Gary-Bobo et al. (2016) develop a multi-stage human capital accumulation model to look at the effects of grade retention in grades 6 to 8 (first three grades of junior high school in France) on 9^{th} (final) grade outcomes. They find a small positive average treatment effect on the treated on test scores, and the effect is higher for lower performer students. However, grade retention decreases the probability of entering 9^{th} grade for all students.

Using rich administrative and survey data from Hungary, I look at grade retention in the 7^{th} grade of primary school, and estimate the differences in educational aspirations in 8^{th} grade between retained and non-retained children conditional on their socioeconomic status. I also look at how the changes in aspirations translate to the type of secondary school track they are in in 10^{th} grade. Though it is important to know how grade retention affects track choice and educational attainment, we know little about the mechanism through which retaining a student leads to these outcomes. Are these children simply not able to catch up, which leads to lower educational outcomes? Or they also lose their confidence and set lower goals which is why their educational outcomes will be worse than those of non-repeaters? To answer these questions, we have to look at how children adjust their aspirations after retention. To my best knowledge, existing papers only study the effect of early grade retention on aspirations. Hughes et al. (2013) find that parental expectations about their child's highest education level decrease after retention in the 1^{st} grade of primary school. They find that decreasing parental expectations play a role in the negative effect of retention on 3^{rd} grade performance of children. Cham et al. (2015) look at the effect of grade retention in the primary grades on students' own expectations in 9^{th} grade about finishing high school. They use propensity score weighting to equate the distribution of pre-treatment characteristics of retained and non-retained students. They do not find any effect of grade retention on motivation to finish secondary school in 9^{th} grade. They find, however, that retained students in 9^{th} grade value a high school diploma more, and feel that their teachers and peers are more likely to expect them to graduate.

Aspirations act as reference points that induce motivation through loss aversion (Heath et al., 1999; Page et al., 2007). These reference points are highly influenced by one's social environment and so, through the individual actions motivated by them, may reinforce economic inequalities (Genicot and Ray, 2017, 2020). If aspirations are so important in what people achieve, it is worth looking

at how they are shaped by negative events during one's educational career. My main outcome of interest is the educational aspirations at 8^{th} grade. I proxy socioeconomic status with parental education. In my main specifications, I define low SES as both parents having lower qualification than a high school diploma, and high SES as at least one parent having a high school diploma. Educational hardships might cause a decline in low-SES students' dream education level, while high-SES students might stay focused on the education level they wanted initially to achieve. I find evidence for heterogeneities in the aspirations and the paths of students after retention.

Low-SES repeaters are 15.6 percentage points more likely to aspire for at most a secondary vocational certificate in 8^{th} grade than non-repeaters, while the share of high-SES wanting to achieve this level does not differ by retained status, controlling for 6^{th} grade aspirations and other characteristics. While already in 6^{th} grade 24.8 percent of low-SES children want to achieve at most a secondary vocational certificate compared to 3.8 percent of high-SES, the SES gap is even larger for retained students. I also look at the aspiration for tertiary education. Here, low-SES repeaters do not change their (already quite low) tertiary aspirations, but for high-SES repeaters, the difference is -7.9 percentage points. Therefore, the SES-gap in tertiary aspirations is actually lower for retained than non-retained students. When looking at the effect of retention on the aspired years of education by pre-retention aspirations, I find that the compensatory advantage of high SES is highest at high levels of initial aspirations: while low-SES children decrease their aspirations significantly, high-SES with initially high aspirations stay on the same track.

In Hungary, a student can be retained if she fails at least one subject in school or if she was missing from school for a significant amount of time and could not pass a grading exam. When controlling for characteristics that might have led to the student failing a subject or missing from school, high-SES children are still less likely to get retained than low-SES children. 6^{th} grade midterm mathematics performance in school seems to be one factor that, if the student performs poorly, affects students' 7^{th} grade retention similarly, regardless of socioeconomic status. When splitting the sample by 6^{th} year midterm mathematics grades, the compensatory advantage disappears for the lowest-performing students but it is there for higher performing ones, where retention is much rarer. Using a smaller, administrative database and controlling for factors that are good proxies for missing from school because of health issues – visits to the general practitioner and days spent in a hospital in 7^{th} grade –, the results stay similar, though they become much weaker because of the low sample size. Finally, I look at the differences by socioeconomic status within these groups in transitioning to a secondary school track that gives access to tertiary education. The patterns here are similar to the aspirations: low-performing students are 11 percentage points less

likely to get into this track if retained, but there is no compensatory advantage of high SES among these students. There is, on the other hand, a compensatory advantage for the higher performer students: low-SES repeaters are 6-24 percentage points less likely to attend this institution, while high-SES repeaters are not less likely to attend than non-repeaters. When controlling for health characteristics in the administrative database, on the other hand, all coefficients become smaller and insignificant, and the compensatory advantage at each mathematics performance level disappears. We have to treat these results with caution, though, because the numbers of repeaters in these samples are very small.

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 rich, administrative and survey, educational data of all Hungarian students in the 6^{th}, 8^{th}, and 10^{th} grades, between 2008 and 2017. For learning more about the reasons behind grade retention I use an administrative dataset that covers half of the Hungarian population of ages 5-74 between 2003 and 2017. This dataset contains demographic and labor market data, along with 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 that covers the period between 2008 and 2017. The National ABC is a standardized mathematics and reading comprehension test that all Hungarian students in 6^{th}, 8^{th}, and 10^{th} grade of public education take. The test has been conducted yearly since the 2005/2006 schoolyear. Compulsory education in Hungary for students in the sample 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 have to write the test, except those with autism and intellectual disabilities.The test is centralized, administered by the Education Authority, and aims at measuring the problem solving skills of students in mathematics and reading comprehension, rather than knowledge of school material. Students take the test in their own school, at the end of May. From 2008, students are identified by a unique identifier, the so-called OM code,[1] so 6^{th}, 8^{th}, and 10^{th} grade data of the same students can be linked. The National ABC database contains the standardized test scores and various background characteristics of students and their families from a background questionnaire.

[1] OM is the abbreviation for the Ministry of Education *(Oktatási Minisztérium)*

2.1.1 Background questionnaire

The tests are complemented by three background questionnaires on student, institution, and, if a school has multiple branches, branch level. Students complete the student questionnaire on paper at home, and the headmaster of the institution, and, in case of multiple branches, the head teacher of the branch completes the institution and the branch level questionnaires online. Completing either questionnaire is voluntary. The student questionnaire contains 47 questions, and students take it home and fill it in with the help of their parents. I only use variables from the student questionnaire, which contains variables regarding the student's academic progress (last year GPA, last midterm GPA and grades 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 rich 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 the Admin3 contains demographic data, labor market status, income, job characteristics, social transfers, and education status of each person in the dataset, on a monthly level. This main dataset is then linked by an individual ID 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.

2.2.1 Health care data

For better predicting retention I use administrative health care data from the Admin3 that comes from the National Health Insurance Fund. 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

For the analysis I have to define the variables of interest. These are the treatment variable – repeating 7^{th} grade, educational aspirations, and socioeconomic status. For the first one, children state in the questionnaire whether they had to repeat a grade once or multiple times in different parts of their school career. I use this variable to control for whether someone repeated a grade up to the 6^{th} grade. As for the treatment variable, I simply treat students as having repeated between 6^{th} and 8^{th} grades if more than two years passed between their first 6^{th} grade test and their first 8^{th} grade test. If there are more instances of the same person in the 6^{th} grade database, I regard these instances as repetitions of 6^{th} grade. From the number of 6^{th} grade observations and the years passed between the first 6^{th} and 8^{th} grade occurrences, I can deduce if a child repeated the 7^{th} grade. Since children take the National ABC test in May, by May they already suspect if they will have to repeat the grade, so 6^{th} grade aspirations partly reflect the effect of later retention. Due to this potential effect on pre-treatment variables, I use 7^{th} grade retention as the treatment and I exclude those who were retained in 6^{th} grade. There is a practice among children who apply for 6-year academic secondary schools in 6^{th} grade but do not get admitted: they complete 7^{th} grade in their primary school, apply again, and if admitted, repeat 7^{th} grade but in the new institution they were admitted to. I set the treatment to be 0 for these children (if they have not repeated the 6^{th} grade either), because their grade repetition is of a very different nature than what I am interested in.

In the questionnaire, children choose the education level they want to reach from a list of qualifications from primary school to a doctoral degree. I create two variables from this data. The first one 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 8 years, a vocational degree is 11 years, and so on (see the construction of the variable in more detail in the Data Appendix section A.4.2).

The last variable that is needed to define is socioeconomic status (SES). For the main specifications, I use a categorical SES variable. I proxy socioeconomic status with parental education, which is also asked in the survey in fine categories: from unfinished primary education to university degree. I create two categories: low-SES is a child whose parents do not have a high school diploma, while high SES means at least one parent has a high school diploma. I chose this distinction after Falk et al. (2019), however, they also use information on family income which unfortunately I do not have. Also, since there are very few repeaters among children of tertiary educated parents, drawing

the line at a lower parental education level gives enough repeaters in both SES categories. In a robustness check I use a three-category SES variable, where high SES is further decomposed into two categories: medium SES here means at least one parent with at least a high school diploma, and high SES means at least one parent with a tertiary degree. In another robustness check, I use a continuous variable, where I apply the same rule to convert the categories to years of education as for the educational aspirations. For the continuous SES variable I take the parent with the highest education level.

2.3.1 Grade retention rules in Hungary

Grade retention in Hungary is regulated by the 2011 Public Education Law.[2] In the first grade of primary school, parents can request grade retention for their child even if the child fulfilled all educational requirements. The school principal has to approve this request. From the second grade and above, grade retention has two sources: first, if the child receives an insufficient grade in at least one subject, she can take a grading exam in those subjects at the end of summer. If she fails, she has to repeat the entire grade. Teachers have a say in the decision about whether to fail someone in a subject or make her do some extra coursework in order to pass the subject without having to take the grading exam. The second source is absenteeism. If the child missed at least 250, she has to take equivalency tests in the subjects in which she cannot receive a final grade at the end of the school year. She has to pass all tests to be able to proceed to the next grade. If the child has at least 250 unjustified absences, she has to repeat the entire grade, without the possibility to take the equivalency test. Retention is also automatic if the child receives an insufficient grade from at least 3 subjects. In practice, grade retention in Hungary is quite rare. Figure 1 shows the percentage of retained students in the upper grades of primary school by year. Usually 2 to 3 percent of students are retained in these grades, and retention in the final grade of primary school is even rarer: less than 1 percent in most years. Grade retention in the lower grades of primary school is not very common either, except in the first grade, where around 3.3-4.5 percent of children were retained in my sample.

There is another reason why someone can be retained: if the family spends one or more years abroad, the child has to prove that she learned everything her peers learned during the schoolyear by passing a grading exam. Often the skills and knowledge the child learns abroad are very different from what the Hungarian school system requires and the child cannot pass the grading exam, leading to having to repeat a grade (see the experiences of returning children in Árendás et al.,

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

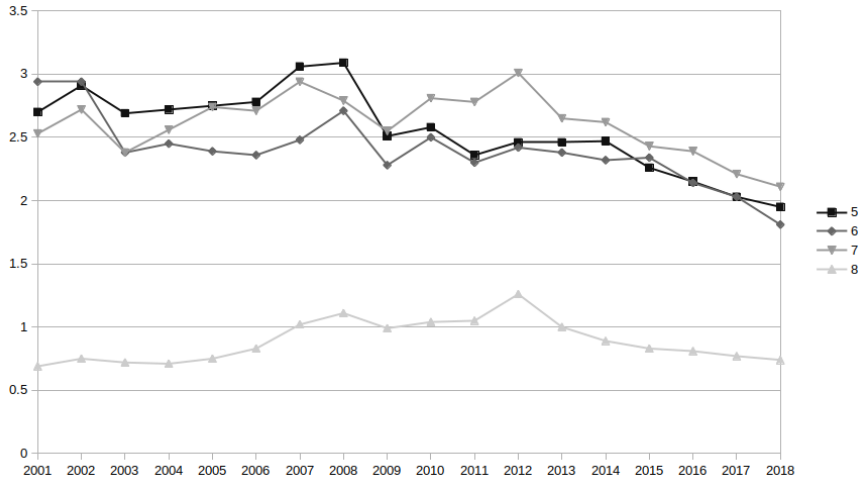

Figure 1: Percentage of retained students in higher grades of primary school

Source: KIR-STAT

2022). Retention because of not meeting the Hungarian requirements is similarly stigmatized as retention because of bad performance in the Hungarian schools, so in the end these children also face a negative shock they have to cope with. 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 for better living conditions (Árendás et al., 2022).

3 Descriptive statistics

3.1 Aspirations

Table 1 shows the educational aspirations in 6th grade by parental education (socioeconomic status). The sample of the table includes children who did not have to repeat any grade until the 6^{th} grade[3]. Low-SES children are significantly more likely to aspire for at most vocational education already in 6^{th} grade than high-SES children (24.78% compared to 3.8%). They are also significantly less likely to aspire for tertiary education than high-SES children (25.12% compared to 69%). While the highest share of low-SES children (50.10%) aspire for a high-school diploma or a post-secondary non-tertiary qualification, for high-SES children the most popular education level is tertiary.

Figure 2 shows how educational aspirations change from 6^{th} to 8^{th} grade by SES and repeater

[3]In fact, low-SES children already have a much higher chance to repeat in the lower grades of primary school than high-SES children (7.5 percent vs. 1.5 percent in my sample). This means that the low-SES children who are selected into my sample were better able to meet school requirements and catch up with their peers.

	Socio-economic status	
Aspirations in 6th grade	Low	High
At most vocational	24.78	3.80
High school diploma or post-secondary non-tertiary	50.10	27.20
College or university	25.12	69.00

Table 1: Educational aspirations of 6^{th} grade children by socioeconomic status

Note: Column percentages. Sample excludes children who have already repeated until 6th grade.

status. I converted aspirations into into years of education for easier interpretation. On average, all non-repeaters increase their aspirations by a little (see Appendix Table 16 for more detail). However, there are large differences between repeaters, especially at higher initial aspirations. For low-SES children, the aspiration gap between repeaters and non-repeaters increases with initial aspirations, while for high-SES children it stays pretty constant, at around a year. Appendix Table 15 shows the number of observations and the share of repeaters in 7^{th} grade in each aspiration category– SES cell.

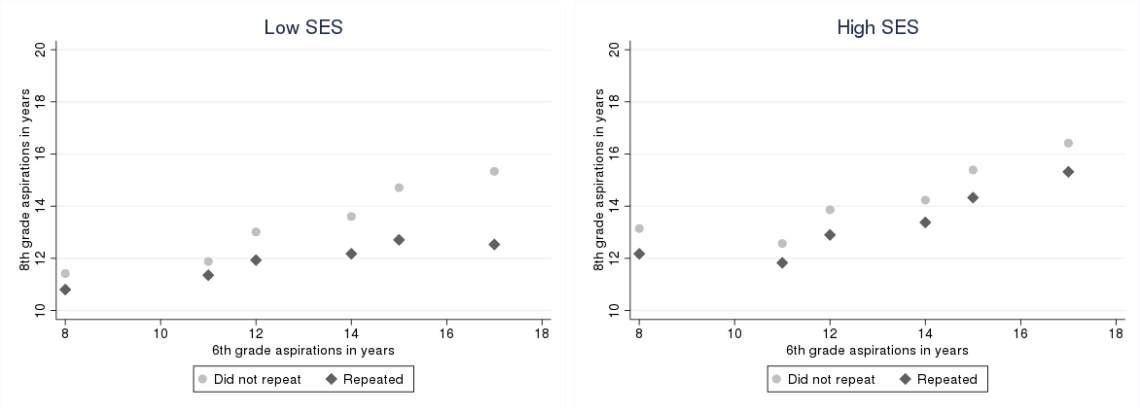

Figure 2: Change in educational aspirations between 6th and 8th grade by SES and repeating

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 sample of the figures exclude children who were in an academic secondary school in 6th grade. It also excludes children who have already repeated until 7^{th} grade.

3.2 Retention

Table 2 shows the probability of children having to repeat 7^{th} grade by socioeconomic stastatustus, conditional on not having repeated until 7^{th} grade. Low-SES children are more likely to get retained: 1.24 percent of them repeat 7^{th} grade, compared to 0.33 percent of high-SES children. When we

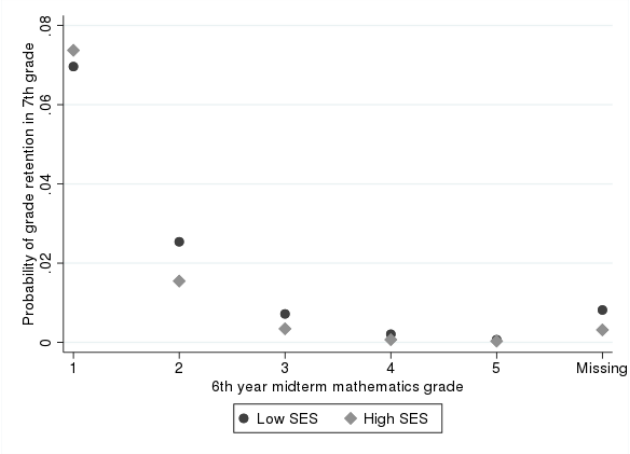

Figure 3: Probability of repeating 7th grade by 6th year midterm mathematics grade

Note: The sample excludes children who have already repeated until 7th grade.

look at their 6^{th} year midterm grades in a core subject – mathematics – in Figure 3, we see that except for the lowest mark, low-SES children have a higher chance to repeat in each category. Among those who failed in mathematics in the first term of 6^{th} grade, high-SES are more likely to repeat 7^{th} grade than low-SES. This might be partly because low-SES in this group are more likely to get retained *already in* 6^{th} *grade*, so excluding 6^{th} grade repeaters introduces a selection into the sample. I will talk about this selection problem and potential solutions later in this section and in Section 7.

	Socio-economic status	
Grade retention in 7th grade	Low	High
Did not repeat	98.76	99.67
Repeated	1.24	0.33

Table 2: Repeating 7th grade grade by socioeconomic status

Note: Column percentages. The sample excludes children who have already repeated until 7th grade.

Since students from different socioeconomic backgrounds have a different chance to repeat 7^{th} grade at similar 6^{th} grade school performance, it is worth looking at repeaters and non-repeaters separately by SES to find out who the repeaters are in each group and what are the predictors of their retention. To be able to compare repeaters and non-repeaters by a wider range of characteristics, for this comparison I use the administrative database. Table 3 is a summary table of the characteristics of repeaters and non-repeaters by SES. Since the size of the administrative database is only half of the National ABC, I have around 500 low-SES repeaters and a bit over 200 high-SES repeaters

in the sample, compared to the 55000 low-SES non-repeaters and 112000 high-SES non-repeaters. The variables I am comparing high-and low-SES children are factors that can affect retention: 6^{th} grade test scores and midterm grades in different subjects, whether their parents lived together in 6^{th} grade and whether they potentially separated between 6^{th} and 8^{th} grades (measured by the child living with both parents in 6^{th} grade but living with only one of them in 8^{th} grade), parental labor market status and its changes between 6^{th} and 8^{th} grades, visits to the general practitioner, and days spent in hospital. In general, both low- and high-SES repeaters perform worse in 6^{th} grade than non-repeaters, however, the average performance of high-SES children is better in all subjects and in both standardized tests compared to low-SES children. There are some differences regarding the family structure: high-SES repeaters are much more likely have parents not living together than low-SES repeaters, suggesting that being raised by a single parent might put a higher risk of retention on them. The share of separated parents are higher among repeaters in both SES groups. Another factor that might be a more frequent cause for retention for high-SES is being hospitalized: high-SES repeaters spend on average twice as much in hospital than low-SES repeaters (1.54 vs 0.75 days), while the difference is quite small for non-repeaters (0.24 vs 0.28 days).

However, when we control for these factors when regressing 7^{th} grade retention on socioeconomic status, high-SES children still seem to have a lower chance to repeat than low-SES. The first column of Table 4 shows the raw difference between the repeating probability of high- and low-SES children. Including the characteristics studied above, the gap between high- and low-SES children decreases significantly, but the difference is still there, especially when including school fixed effects. Within the same school, high-SES children are less likely to get retained, even when controlling for a rich set of characteristics that affect retention. Column 4, however, shows that at the lowest 6^{th} year midterm mathematics grade, high-SES are equally likely to get retained than low-SES. Table 5 shows the predictors of repeating across 6^{th} year midterm grade groups. As the first column shows, SES does not affect retention in the lowest mathematics grade group, but it does affect retention at higher mathematics grades. For the lowest-performing students in mathematics, marks in other subjects and test scores do not seem to affect retention, neither family structure nor parental labor market status variables, or health characteristics. The only factor that still affects retention is the 6^{th} year behavior grade of the child, which is a measure of how well the child behaves at school according to her teachers. In this group, retention is 6.5 percent. The second and third column shows those children who got a 2 and at least a 3 in mathematics at 6^{th} grade midterm. The share of retained are 1.9 and 0.18 percent in these groups, and here, high-SES are ceteris paribus still less likely to repeat. For these groups, bad marks in another core subject, Hungarian literature,

	Low - Did not repeat	Low - Repeated	High - Did not repeat	High - Repeated
6th grade mathematics test	1441.95	1318.58	1563.89	1402.74
score	(173.12)	(146.07)	(174.06)	(167.27)
6th grade reading test score	1430.68	1296.33	1562.56	1379.33
	(174.38)	(148.46)	(174.04)	(179.21)
6th year midterm mathematics	3.19	2.06	3.91	2.36
grade	(1.02)	(0.73)	(0.96)	(0.99)
6th year midterm literature	3.63	2.48	4.28	2.88
grade	(0.98)	(0.81)	(0.82)	(1.04)
6th year midterm Hungarian	3.43	2.39	4.06	2.84
grammar grade	(0.97)	(0.77)	(0.88)	(1.01)
Parents lived together in 6th	0.77	0.67	0.79	0.50
grade	(0.42)	(0.47)	(0.41)	(0.50)
Parents separated from 6th to	0.05	0.11	0.04	0.12
8th grade	(0.22)	(0.31)	(0.21)	(0.33)
Mother does not work in 6th	0.37	0.47	0.15	0.23
grade	(0.48)	(0.50)	(0.36)	(0.42)
Mother has a permanent job in	0.50	0.40	0.74	0.62
6th grade	(0.50)	(0.49)	(0.44)	(0.49)
Father does not work in 6th	0.15	0.24	0.06	0.12
grade	(0.36)	(0.43)	(0.24)	(0.33)
Father has a permanent job in	0.64	0.54	0.72	0.65
6th grade	(0.48)	(0.50)	(0.45)	(0.48)
Mother stopped working between	0.07	0.09	0.04	0.09
6th and 8th grades	(0.26)	(0.29)	(0.19)	(0.28)
Father stopped working between	0.06	0.08	0.03	0.05
6th and 8th grades	(0.23)	(0.26)	(0.16)	(0.21)
Number of GP visits in 7th	4.99	7.57	4.05	6.31
grade	(4.82)	(6.74)	(4.07)	(6.87)
Days spent in hospital in 7th	0.28	0.75	0.24	1.54
grade	(1.76)	(4.82)	(1.70)	(9.02)
Observations	54812	493	112171	216

Table 3: Summary statistics of low- and high-SES 7th grade repeaters

Note: Rows 6-13 show shares of students with that characteristic, the rest of the rows show average levels. 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.

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. These groups are therefore more heterogeneous in terms of the reason for repeating.

In the empirical analysis, first I look at the compensatory advantage of high socioeconomic status

	Repeating 7th grade			
	(1)	(2)	(3)	(4)
High SES	-0.00783***	-0.000679*	-0.00149***	0.0103
	(0.00039)	(0.00037)	(0.00040)	(0.0099)
6th year midterm mathematics grade, baseline: 1 (fail)				
2		-0.0248***	-0.0247***	-0.0195***
		(0.0049)	(0.0049)	(0.0058)
3		-0.0304***	-0.0302***	-0.0276***
		(0.0048)	(0.0048)	(0.0057)
4		-0.0305***	-0.0301***	-0.0278***
		(0.0048)	(0.0049)	(0.0057)
5		-0.0294***	-0.0287***	-0.0263***
		(0.0048)	(0.0049)	(0.0057)
High SES × 2				-0.0170*
				(0.0100)
High SES × 3				-0.0109
				(0.0099)
High SES × 4				-0.0106
				(0.0099)
High SES × 5				-0.0107
				(0.0099)
Constant	0.0104***	0.0942***	0.0987***	0.0950***
	(0.00039)	(0.023)	(0.023)	(0.023)
Observations	243989	216529	216529	216529
Controls	no	yes	yes	yes
Year fixed effect	no	yes	yes	yes
School fixed effect	no	no	yes	yes

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

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

Table 4: Regression of retention on SES and various characteristics.

Note: The controls include variables presented in Table 3. For all variables, dummies are used for missing values.

on aspirations on the whole sample, then I divide the sample by 6^{th} year mathematics grades to see if the compensatory advantage is different across these groups. Children with the lowest mark in mathematics most likely repeat because they perform badly at school, but the other two groups are more heterogeneous and have to be treated with caution. Mathematics is a fairly objective subject where students' knowledge can be easily assessed. However, other subjects give more leeway to the subjective judgment of teachers, and the advantages of higher socioeconomic background might show in other ways in these subjects: e.g. with similar lexical knowledge high-SES children might still be better in essay writing. These reasons would lead to high-SES students being retained in 7^{th} grade less likely than low-SES students. It is also possible that after having bad marks in 6^{th} grade but not having to repeat in the end, high-SES students start to catch up faster than

low-SES students, so the lower chance of high-SES repeating 7^{th} grade might already capture some compensatory advantage in catching up after a risk of retention. Therefore, if I find a compensatory advantage of high SES in aspirations after retention, it is likely a lower bound of what I would find for the first retention shock and if high- and low-SES children of similar observable characteristics had equal chances to repeat. In the 3-5 mathematics grade group the advantage of high-SES in retention is very low and significant only at the 10 percent level. The factors that affect the group who had a 2 in mathematics affect this group's retention, as well, but retention in this group is extremely rare.

3.3 8^{th} grade aspirations and secondary school tracks

Though unfortunately I cannot follow the students for long in the database, so I cannot look at how 8^{th} grade aspirations relate to later life outcomes, Figure 4 shows that they are good predictors of the type of secondary education students attend in 10^{th} grade. 66 percent of students with tertiary aspirations are in an academic secondary school at 10^{th} grade, and around 32 percent of them in a technical secondary school. Only 1.3 percent ends up in a vocational school. At the other end, 90 percent of students aspiring for at most a vocational certificate go to a vocational secondary school, 8 percent of them to a technical secondary, and around 2 percent to an academic secondary school. We have to note that students take the National ABC at the end of May, while 8^{th} 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 8^{th} grade aspirations and 10^{th} grade secondary school outcomes separately, as they are highly correlated. However, the high correlation also validates the aspirations measure, strengthening the interpretation that aspirations reflect children's true preferences about their future education.

3.4 Missing data

As completing the student background questionnaire is voluntary, there is a selection bias in my sample if completion is nonrandom. In the analysis I include observations with missing values by using dummies, however, I cannot apply this technique for socioeconomic status and the outcome variables. 6 percent of the sample has missing parental education data, so these observations have to be excluded from the analyses. Looking at 6^{th} grade mathematics test scores, the average in this sample with missing parental education is 1447 points, compared to the 1407 of low-SES and 1546 of high-SES children. In this sample the share of children I classify as repeaters 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 8^{th}

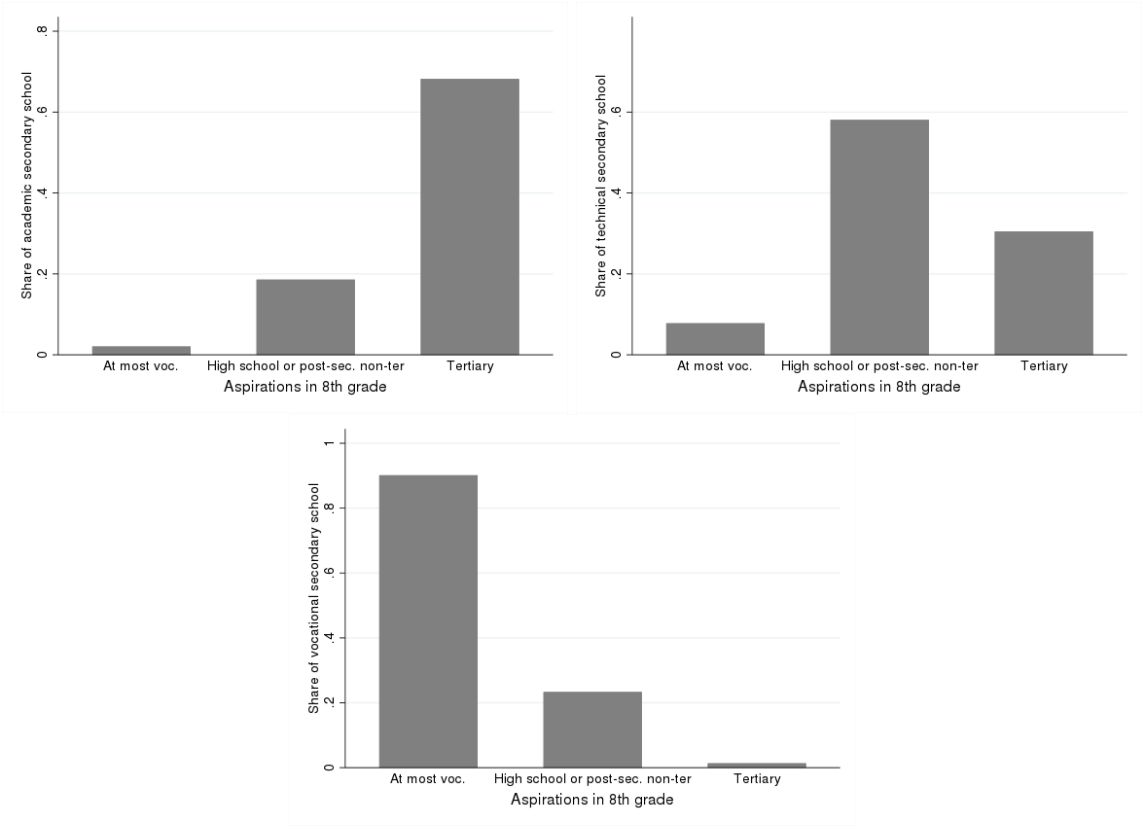

Figure 4: Probability of attending a type of secondary school by 8^{th} grade aspirations

Note: Percent of students with a particular 8th grade aspiration in different secondary education tracks in 10th grade.

grade educational aspirations: on average these children aspire for 13.7 years of education, compared to 13.3 years for low-SES and 15.4 for high-SES children. Based on these, there is a selection bias in completion towards children who perform better at school, have higher aspirations, and who most likely have higher educated parents. This means that the results I find might underestimate the compensatory advantage of high socioeconomic status, as I lose many observations of low-SES children.

4 Baseline empirical analysis

4.1 8^{th} grade aspirations

First I look at how grade retention in 7^{th} grade relates to 8^{th} grade aspirations and type of secondary education by socioeconomic status. In my main specifications, I estimate the following equation:

$$y_{i,s,t} = \alpha_0 + \alpha_1 \text{Repeated}_i + \alpha_2 \text{High SES}_i + \alpha_3 \text{Repeated}_i \times \text{High SES}_i + \beta_1 X_i + \beta_2 X_{6,i,t} + \gamma_s + \eta_t + \epsilon_{i,s,t} \quad (1)$$

Here, $y_{i,s,t}$ is one of two binary outcomes: one is 1 if student i from school s who was in 6^{th} grade in year t wants to reach at most vocational education in 8^{th} grade, and 0 if she has higher aspirations. The second outcome is 1 if the student wants to obtain a tertiary certificate in 8^{th} grade and 0 if she has lower aspirations. The main coefficients of interest are α_1 and α_3: α_1 shows how the aspirations of a low-SES child change if she has to repeat 7^{th} grade, and α_3 shows if this change is different for high-SES children. X_i are time-invariant characteristics of the child, such as gender and whether she has special education needs, and $X_{6,i,t}$ are other pre-treatment, 6^{th} grade characteristics, like 6^{th} grade GPA, mathematics and reading test scores, and the parents' labor market status. γ_s are 6^{th} grade school fixed effects, and η_t are 6^{th} grade year fixed effects.

Table 6 shows the effect of grade retention on the probability of aspiring for at most vocational education in 8^{th} grade. Column 1 shows the raw difference between repeaters and non-repeaters. Repeaters are 38.7 percentage points more likely to aspire for vocational education in 8^{th} grade than non-repeaters. In column 2 I include socioeconomic status and its interaction with repeating. We can see that low-SES repeaters are 37.8 percentage points more likely to aspire for vocational education than non-repeaters, while high-SES repeaters 21.1 percentage points. The differences remain large even if we control for 6^{th} grade aspirations and other time-invariant and 6^{th} grade characteristics in columns 3 and 4, and include school fixed effects in column 5. When including all controls and school fixed effects, low-SES repeaters are 15.6 percentage points more likely to aspire for vocational education than high-SES repeaters do not seem to change their vocational aspirations significantly.

Table 7 shows the effect of retention in 7^{th} grade on the probability of tertiary aspirations in 8^{th} grade. Repeaters are on average 42.8 percentage points less likely to aspire for tertiary education in 8^{th} grade than non-repeaters. This difference is heterogeneous by SES: in column 2 we can see that low-SES repeaters are 21.5 percentage points less likely to aspire for tertiary education than non-repeaters, but for high-SES repeaters the difference is even higher: -42.9. Although the difference is much higher for high-SES children, still 27.2 percent of repeaters in this group aspire for tertiary education, while for low-SES children this share drops from 25.4 percent to 3.9 percent. When controlling for 6^{th} grade aspirations (column 3) other 6^{th} grade controls (column 4), and school fixed effects (column 5), the difference between low-SES repeaters and non-repeaters decreases sharply, and disappears. For high-SES repeaters the difference remains: they are around 7.9 percentage points less likely to aspire for tertiary education than non-repeaters.

Table 8 explores Figure 2 in a regression framework, namely, whether the aspirations of repeaters and non-repeaters change differently by how high their initial aspirations were. Here I run the

following regression:

$$\begin{aligned} y_{i,s,t} = \alpha_0 &+ \alpha_1 \text{Repeated}_i + \alpha_2 \text{High SES}_i + \alpha_3 \text{Repeated}_i \times \text{High SES}_i \\ &+ \alpha_4 \text{6th grade aspirations}_i + \alpha_5 \text{Repeated}_i \times \text{6th grade aspirations}_i \\ &+ \alpha_6 \text{6th grade aspirations}_i \times \text{High SES}_i \\ &+ \alpha_7 \text{Repeated}_i \times \text{6th grade aspirations}_i \times \text{High SES}_i + \beta_1 X_i + \beta_2 X_{6,i,t} + \gamma_s + \eta_t + \epsilon_{i,s,t} \end{aligned} \tag{2}$$

6^{th} grade aspirations are now measured by the years of education the child wants to achieve. α_5 shows whether children with initially high aspirations degrade their aspirations more after grade retention, and α_7 shows if this decrease is heterogeneous by socioeconomic status. 6^{th} grade aspirations are demeaned, so the baseline of the interactions always show the effects at the average level of initial aspirations. As we can see in column 1 of Table 23, non-repeaters aspire for almost 15 years of education on average, which corresponds to a college or BA-level degree. Repeaters aspire for 2.5 years less, which is nearly equivalent to a high school diploma. When we control for socioeconomic status in column 2, we see that high-SES non-repeaters aspire for almost 2 years more education than low-SES non-repeaters, but the drop after retention in aspirations is higher for high-SES than for low-SES, by a third of a year. The negative coefficient on the interaction can be explained by high-SES children having initially higher aspirations, so they can decrease more on average with retention. Controlling for 6^{th} grade aspirations, the difference between repeaters and non-repeaters by parental education virtually disappears. On average, there is no compensatory advantage of high SES in aspirations. However, when in column 4 I interact retention and parental education with 6^{th} grade aspirations, the average negative effect seems to mask a heterogeneous effect by initial aspirations. Column 5 adds 6^{th} grade controls, while column 6 adds school fixed effects. Column 6 shows that a low-SES child with average 6^{th} grade aspirations (14.5 years) decreases her aspirations by 10 months (0.85 years) in case of having to repeat 7^{th} grade. In contrast, high-SES repeaters with average 6^{th} grade aspirations decrease their aspirations by only 3 months. Retention causes a larger drop the higher the initial aspirations were for low-SES children, by about 2 months (0.163 years) for each year of 6^{th} grade aspirations. However, having high socioeconomic status completely offsets this penalty of higher initial aspirations.

5 Subsamples of repeaters by 6^{th} grade mathematics performance

To be able to separate children with different reasons for repeating, now I present the results of the regressions similar to Equation 2 ran on subsamples based on 6^{th} grade midterm mathematics performance. For these regressions, I also use the National ABC database, because there is already

half of the sample in the Admin3 dataset, and with the very few repeaters I would have too few observations in each of these subsamples. The drawback of using the National ABC is that I cannot control for GP visits and hospitalizations, which, as we saw in Table 5, affect retention in the groups with better 6^{th} grade mathematics performance. Appendix Section A.3.4 presents the regressions below ran on the Admin3 data and including controls for GP visits and days spent in a hospital.

Table 9 shows the results for the lowest performer students in mathematics. The share of repeaters is 7.1 percent in this group, and it is quite similar by SES: 6.9 percent of low-, and 7.4 percent of high-SES children in this group repeat. The first column of Table 9 shows that within this group, repeating does not decrease future aspirations significantly for either low- or high-SES children. High-SES non-repeaters, on the other hand, aspire for about 8.5 months (0.7 years) higher education than low-SES non-repeaters. The second column controls for initial aspirations, and the third one interacts aspirations with socioeconomic status, assuming that socioeconomic status has a differential compensating effect at different levels of aspirations. When conrolling for initial aspirations, we see that in this group, repeaters do decrease their – already low, about 12 years – aspirations but only by around 3 months (0.24 years), and the decrease is only significant on the 10 percent level.

The signs of the coefficients, except for the triple interaction of repeating, high SES, and 6^{th} grade aspirations, are similar to the full-sample ones, though they are smaller in magnitude. Part of the reason for the smaller magnitude is that children with bad mathematics performance already had low aspirations in 6^{th} grade. Combined with the low sample size of this group,[4] the interactions of repeating and SES are not significant. However, the coefficient on the triple interaction is negative, meaning that high-SES children with higher initial aspirations in this group decrease their aspirations more than low-SES children if they repeat. Although the coefficient is insignificant, the opposite sign suggests that in this low-performer group, both low- and high-SES children suffer from retention. socioeconomic status does seem to matter though for the non-repeaters. Low-performer children who in the end did not have to repeat the 7^{th} grade aspire for higher education levels if they are from high SES, and the difference is even higher for children with higher initial aspirations.

Appendix Table 25 shows the results of the same regression ran on the Admin3 data with controls for GP visits and hospitalization in 7^{th} grade as well. The only coefficients that are large and significant are socioeconomic status for non-repeaters – they aspire for 7 months (0.6 years) higher education than low-SES –, and 6^{th} grade aspirations. Low-SES repeaters decrease their aspirations by 3 months (0.24 years), while high-SES by 3.5 months (0.3 years), though none of

[4]The number of repeaters are 281 in the low-SES, and 114 in the high-SES group.

the differences are significant. This sample is very small, so the non-significant results might be because of low sample size.[5] Each extra year of 6^{th} grade aspirations decreases repeaters' 8^{th} grade aspirations, and for high-SES repeaters the decrease is even larger, though these differences are not significant, either. It seems that in this low-performer group, compensatory advantage does not work after retention, leaving both high- and low-SES children with lowered expectations about their educational attainment.

Table 10 shows the results on the group who have just passed mathematics at 6^{th} grade midterm. In this group, 2.2 percent had to repeat 7^{th} grade (2.5 percent of low-SES, and 1.5 of high-SES).[6] Children in this group aspire for 13 years of education in 6^{th} grade on average. Repeaters have around half a year lower aspirations in 8^{th} grade than non-repeaters, and without controlling for aspirations, this difference is similar for high- and low-SES children. The second column controls for 6^{th} grade aspirations, and the third column introduces the triple interaction. In the third column, we see similar patterns to the full sample: there is a compensatory advantage of high socioeconomic status in aspirations for repeaters with average initial aspirations. While low-SES repeaters decrease their aspirations by half a year, for high-SES repeaters this decrease is only a fifth of a year. Each extra year of 6^{th} grade aspirations increases the drop in aspirations after retention by a little bit more than a month. This increase is lower for high-SES children, though not significantly. The differences in the aspirations of high- and low-SES non-repeaters are there in this group, too.

Appendix Table 26 presents the results of the same regression on the Admin3 data, including controls for 7^{th} grade health characteristics.[7] The picture here is similar to that in Table 10, except that while low-SES repeaters with average 6^{th} grade aspirations decrease their 8^{th} grade aspirations by about half a year, high-SES repeaters do not decrease them significantly. While the repeater–non-repeater aspiration gap increases for low-SES children by 1 month per each year of initial aspirations, the gap is constant and not significantly different from zero for high-SES children. In this group there is already significant compensatory advantage of high socioeconomic status.

Finally, Table 11 shows the effect of retention on aspirations for those whose 6^{th} grade midterm mathematics performance was at least average. In this group, retention was extremely rare: 0.23 percent on average, with 0.45 for low- and 0.13 for high-SES students.[8] The average of 6^{th} grade aspirations is 15 years in this group. On average, there does not seem to be a compensatory advantage for repeaters in aspirations in this group either. However, in the third column, where we assume

[5] In the estimation sample there are 117 low-SES repeaters and 55 high-SES repeaters.

[6] This gives 815 low-SES and 301 high-SES repeaters.

[7] In this sample, there are 380 low-SES repeaters and 118 high-SES repeaters.

[8] 311 low-SES and 218 high-SES repeaters.

that SES helps children differently at different initial aspirations, there is a large compensatory advantage at the average aspirations, that increases even more if initial aspirations are higher. Low-SES repeaters decrease their aspirations by a whole year in this group, while for high-SES repeaters the decrease is only 4.3 months (0.36 years). Every extra year of 6^{th} grade aspirations increase the repeater–non-repeater gap by 2 months (0.16 years) for low-SES, but for high-SES, the gap is constant.

Appendix Table 27 controls for GP visits and hospital stays in the Admin3 database. Controlling for these characteristics (and having half the sample size of the National ABC database), the compensatory advantage of high SES in this group becomes insignificant, though still large and positive. The number of repeaters in this sample is also quite small (147 low-SES repeaters and 110 high-SES repeaters), so it is possible that there is still a compensatory advantage among children repeating for similar reasons, but it is more heterogeneous than in the group of children just passing mathematics in 6^{th} grade.

It is likely, therefore, that the compensatory advantage I find on the overall sample is driven by repeaters in the higher performer groups, who are more likely repeating because of health shocks, or other, unobservable reasons. It is also possible that some children in this group repeat 7^{th} grade because the family spent a year abroad; a factor that I do not observe in the dataset. However, as Appendix Table 18 shows, even in the highest performer group, repeaters' 6^{th} year test scores and grades are worse than non-repeaters', and they visited the GP more times and spent more days in a hospital than non-repeaters. Based on these it is unlikely that this group is mostly composed of children who spent a year abroad with their family. Even if there were many such children, retention because of not being able to meet the Hungarian requirements after returning from abroad is also stigmatized (Árendás et al., 2022), so it is not entirely wrong to treat this type of retention as a similar negative shock to other types of retention.

6 10^{th} grade outcomes

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, so one that ends with a high school diploma. In Hungary in my sample period these were the academic secondary schools and the technical secondary schools. I estimate the same equation as Equation 2, except that the outcome variable here is a dummy indicating whether the child is in a technical or an academic secondary school in 10^{th} grade or in a vocational school that does not provide access to tertiary education. I look at the results in the three subsamples that I analysed in the previous section: those who

failed mathematics in 6^{th} grade midterm, those who only passed, and those who had higher marks. Tables 12-14 show the results of the regressions by subsample.

For the lowest performers in mathematics, the patterns are similar to the aspiration changes – partly because in 8^{th} grade, when stating their aspirations, children already know which secondary school they were admitted to. In Table 12 we see that repeaters are 11 percentage points less likely to attend an institution giving access to tertiary education than non-repeaters, and for each extra year of initial aspirations, the probability of attending this institution decreases by 5.9 percentage points. This decrease is not compensated significantly by SES at either initial aspiration level. High-SES non-repeaters, on the other hand, are more likely to attend this institution at every aspiration level than low-SES non-repeaters. When using the Admin3 database and including GP visits and hospital days in Appendix Table 28, the picture is similar, though because of the even lower sample size, none of the interactions with high SES are significant. Comparing children who repeated for the same reason, repeater low-SES are 19 percentage points less likely to attend a secondary school giving a high school diploma, and each extra initial year of aspirations widens this gap by 14 percentage points. The chances of high-SES repeaters at the average aspiration level to attend this institution are at least as much lower than low-SES repeaters' compared to non-repeaters, though initial aspirations seem to – though non-significantly – compensate for them. The advantage of high-SES non-repeaters remains significant.

Interestingly, while for children just passing we saw a large compensatory advantage for repeaters at the average aspirations, this compensatory advantage does not carry on strongly to an advantage in attending a secondary school ending with a high school diploma. Low-SES repeaters are 6 percentage points less likely to attend this institution than non-repeaters, though the difference is only significant on the 10 percent level. For high-SES repeaters the measure of the aspiration gap is quite noisy, but for them, the decrease in aspirations is not significantly different from zero. Table 29 controls for GP visits and hospital days, as well. The aspiration gap for low-SES children remain similar, though it becomes insignificant. There seems to be no compensatory advantage of high SES when comparing children repeating for similar reasons; the coefficient on the interaction of high SES and repeating even becomes negative. The number of repeaters here is very small, as well. The advantage of high-SES non-repeaters is strong and significant in this sample, too.

There is a large compensatory advantage of high socioeconomic status in the sample of students performing well in mathematics. Here repeaters are, both statistically and economically significantly, less likely to attend an institution ending with a high school diploma than non-repeaters. Low-SES repeaters are 25 percentage points less likely to attend this institution than non-repeaters,

while for high-SES children, the difference is non-significant. The compensatory advantage is even higher at higher initial aspirations. The higher 6^{th} grade aspirations are, the more likely low-SES non-repeaters are attending a secondary school that gives access to tertiary education, but for high-SES non-repeaters, the chances of attending this institution does not depend so strongly on initial aspirations.

Appendix Table 30 includes GP visits and hospital days and uses the Admin3 database. The compensatory advantage found without the health controls disappears when we compare children who repeated for the same reason. Without controlling for the interactions of initial aspirations, repeaters are 22 percentage points less likely to attend a high school providing access to tertiary education, and this difference is similar for high- and low-SES children. However, when we control for the interactions of initial aspirations, the aspiration gap at the average initial aspiratoins becomes insignificant for both SES groups. Those low-SES repeaters who had higher initial aspirations are 2.5 percentage points more likely to attend this institution per each year of initial aspirations, while for high-SES the chances of attending does not depend on initial aspirations.

7 Discussion

I find that children from different socioeconomic backgrounds cope differently with having to repeat the 7^{th} grade of primary school: all repeaters decrease their aspirations, but the magnitudes are larger for low-SES children. The post-retention SES-gap in aspirations is the highest with initially high aspirations. When looking at subsamples by previous academic performance in a core subject, mathematics, the picture changes sharply. Students with poor prior mathematics performance decrease their aspirations if they get retained in 7^{th} grade, but the decrease is similar for all children, regardless of socioeconomic background. When controlling for factors that proxy missing school because of health issues, the coefficients on compensatory advantage even turn negative, though insignificant. These children are then all less likely to end up in a secondary school giving access to tertiary education, but there is no difference by socioeconomic status here, either. However, in these groups, socioeconomic status has an advantage for the non-repeaters in both aspirations and secondary school track. It is possible therefore that, in line with the compensatory advantage literature, those students who were on the edge of repeating but in the end they did not have to, are able to catch up better at school if they have a better socioeconomic background. However, retention might be, in line with most findings in the retention literature, such a big negative shock that even high socioeconomic status cannot offset it.

On the other hand, I find a compensatory advantage for those children who performed better

previously, in both 8^{th} grade aspirations and secondary track choice. Retention is quite rare in these groups, especially among those with good mathematics performance, and is likely affected by other factors – such as health issues, separation of parents, or spending a year abroad with the family – than only bad performance. Except for the lowest-performer students, high-SES children always have a lower chance to repeat conditional on all observable factors that can affect retention. It is likely that, since I excluded children who repeated the 6^{th} grade, I introduced an extra selection into the treatment: those children who almost had to repeat 6^{th} grade but they did not have to in the end, might differ by SES in terms of how they catch up with the school workload and how much effort they put into avoiding further risk of retention. The compensatory advantage I find in these groups might be a lower bound of the compensatory advantage I could estimate had I not introduced a selection by excluding repeaters in the 6^{th} grade, so if high-SES children had the same chance to repeat 7^{th} grade as low-SES. In these groups, when controlling for GP visits and hospital days, the compensatory advantage in aspirations remain but the compensatory advantage in attending a secondary school that provides access to tertiary education diminishes, though the sample sizes here are very small.

What could be the mechanisms behind the compensatory advantage? Families try to avoid downward mobility by parents pushing their children towards education levels at least as high as their own. After a negative shock, such as retention, parents try to still push their children towards better education, but high-SES parents have more resources to do so, and they also start from a higher reference point (Bernardi, 2014). Another reason why I see different responses in aspirations by high- and low-SES children might be that their parents are not equally involved in completing the questionnaire. Müller (2021) finds that the aspiration gap between high- and low-SES children is smaller if parents are not informed of children's answers to the aspiration question. The result is driven by high-SES children who aspire for tertiary education when their parents see their answers but for lower education when they do not. It is possible that, while in 6^{th} grade, both high- and low-SES parents help their children completing the questionnaire, in 8^{th} grade, children mostly do it on their own, but the parents of retained students still monitor their children's answers, leading to still high goals for high-SES, but reduced goals for low-SES children.

	Repeated 7th grade		
	(1)	(2)	(3)
	1	2	3-5
High SES	0.00432	-0.00888***	-0.000557*
	(0.018)	(0.0018)	(0.00029)
6th grade mathematics test	0.0000177	-0.0000213***	-0.00000279***
score	(0.000055)	(0.0000070)	(0.00000088)
6th grade reading test score	-0.000100	-0.0000236***	-0.000000833
	(0.000067)	(0.0000067)	(0.0000010)
6th year midterm literature	-0.0160	-0.00484***	-0.00120***
grade	(0.011)	(0.0013)	(0.00026)
6th year midterm Hungarian	0.000548	-0.00209	0.000329
grammar grade	(0.011)	(0.0013)	(0.00021)
6th year midterm effort grade	-0.00467	-0.00925***	-0.00101***
	(0.012)	(0.0015)	(0.00030)
6th year midterm behavior	-0.0281***	-0.00928***	-0.00153***
grade	(0.0093)	(0.0012)	(0.00026)
Parents lived together in 6th	-0.0195	-0.00813***	-0.000744**
grade=1	(0.016)	(0.0019)	(0.00029)
Parents separated from 6th to	0.0515	0.0170***	0.00143**
8th grade=1	(0.035)	(0.0045)	(0.00066)
Mother does not work in 6th	-0.00340	-0.000228	0.000618
grade=1	(0.022)	(0.0027)	(0.00043)
Mother has a permanent job in	-0.0136	0.000231	-0.000165
6th grade=1	(0.024)	(0.0027)	(0.00033)
Father does not work in 6th	0.0377	0.00388	0.000636
grade=1	(0.024)	(0.0031)	(0.00053)
Father has a permanent job in	0.0108	-0.000320	-0.000213
6th grade=1	(0.020)	(0.0021)	(0.00025)
Mother stopped working between	-0.00644	0.00578	0.00112
6th and 8th grades=1	(0.030)	(0.0037)	(0.00069)
Father stopped working between	-0.0106	-0.00619*	0.000469
6th and 8th grades=1	(0.035)	(0.0035)	(0.00076)
Number of GP visits in 7th	0.00249	0.00160***	0.000175***
grade	(0.0016)	(0.00025)	(0.000046)
Days spent in hospital in 7th	-0.00275	0.00220**	0.00103***
grade	(0.0025)	(0.0011)	(0.00038)
Constant	0.300***	0.163***	0.0223***
	(0.10)	(0.013)	(0.0019)
Observations	3303	33271	180700
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 5: Regression of retention on SES and various characteristics by 6th year midterm mathematics grade.

Note: The controls include variables presented in Table 3. For all variables, dummies are used for missing values. Mathematics grades 3-5 are pooled together because of a low number of repeaters in these groups.

	Aspirations in 8th grade: at most vocational school				
	(1)	(2)	(3)	(4)	(5)
Repeated	0.387***	0.378***	0.254***	0.153***	0.156***
	(0.011)	(0.013)	(0.013)	(0.012)	(0.012)
High SES		-0.194***	-0.107***	-0.0821***	-0.0669***
		(0.0024)	(0.0016)	(0.0013)	(0.0013)
Repeated × High SES		-0.167***	-0.140***	-0.138***	-0.139***
		(0.021)	(0.019)	(0.018)	(0.018)
6th grade aspirations, baseline: at most vocational					
High school diploma or			-0.376***	-0.289***	-0.275***
post-secondary non-tertiary			(0.0031)	(0.0030)	(0.0029)
College or university			-0.431***	-0.294***	-0.273***
			(0.0033)	(0.0031)	(0.0030)
Constant	0.0991***	0.223***	0.515***	0.878***	0.813***
	(0.0019)	(0.0026)	(0.0036)	(0.032)	(0.032)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

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

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

Table 6: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma and high SES means at least one parent has a high school diploma. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	Aspirations in 8th grade: tertiary				
	(1)	(2)	(3)	(4)	(5)
Repeated	-0.428***	-0.215***	-0.0718***	0.00490	0.00200
	(0.0072)	(0.0053)	(0.0058)	(0.0056)	(0.0056)
High SES		0.447***	0.229***	0.177***	0.147***
		(0.0032)	(0.0024)	(0.0022)	(0.0018)
Repeated × High SES		-0.214***	-0.133***	-0.0849***	-0.0789***
		(0.018)	(0.016)	(0.015)	(0.015)
6th grade aspirations, baseline: at most vocational					
High school diploma or			0.141***	0.0440***	0.0365***
post-secondary non-tertiary			(0.0019)	(0.0019)	(0.0019)
College or university			0.646***	0.400***	0.373***
			(0.0023)	(0.0026)	(0.0026)
Constant	0.541***	0.254***	0.00958***	-0.536***	-0.425***
	(0.0043)	(0.0021)	(0.0020)	(0.025)	(0.026)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

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

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

Table 7: Probability of 8^{th} grade aspirations being tertiary education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma and high SES means at least one parent has a high school diploma. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	8th grade educational aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-2.462***	-1.734***	-0.943***	-1.691***	-0.832***	-0.850***
	(0.043)	(0.043)	(0.042)	(0.078)	(0.072)	(0.070)
High SES		1.883***	0.827***	0.840***	0.559***	0.446***
		(0.016)	(0.0094)	(0.0094)	(0.0075)	(0.0064)
Repeated ×		-0.345***	-0.0184	0.687***	0.601***	0.599***
High SES		(0.097)	(0.082)	(0.12)	(0.11)	(0.11)
6th grade			0.555***	0.525***	0.341***	0.320***
aspirations in years			(0.0019)	(0.0027)	(0.0027)	(0.0026)
Repeated ×				-0.279***	-0.163***	-0.163***
6th grade aspirations in years				(0.023)	(0.021)	(0.021)
High SES ×				0.0549***	0.0328***	0.0193***
6th grade aspirations in years				(0.0033)	(0.0030)	(0.0029)
Repeated ×				0.202***	0.139***	0.145***
High SES × 6th grade aspirations in years				(0.042)	(0.039)	(0.038)
Constant	14.71***	13.50***	14.12***	14.08***	11.06***	11.34***
	(0.019)	(0.013)	(0.0079)	(0.0085)	(0.16)	(0.15)
Observations	391412	391412	391412	391412	391412	391412
6th 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 class level. Clustered standard errors in parentheses.

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

Table 8: Educational aspirations in 8^{th} grade in years

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma and high SES means at least one parent has a high school diploma. 6th grade aspirations are demeaned. In column 4, 6th grade aspirations are added as controls. In column 5, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 6 adds 6th grade school fixed effects.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.198	-0.222*	-0.239*
	(0.14)	(0.13)	(0.13)
High SES	0.695***	0.515***	0.460***
	(0.069)	(0.067)	(0.067)
Repeated ×	-0.0719	-0.0193	0.0945
High SES	(0.25)	(0.24)	(0.25)
6th grade		0.254***	0.228***
aspirations in years		(0.017)	(0.020)
Repeated ×			-0.0385
6th grade aspirations in years			(0.066)
High SES ×			0.111***
6th grade aspirations in years			(0.034)
Repeated ×			-0.122
High SES × 6th grade aspirations in years			(0.12)
Observations	6338	6338	6338
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 9: 8^{th} grade aspirations for those failing mathematics at 6^{th} grade midterm

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.474***	-0.431***	-0.542***
	(0.050)	(0.050)	(0.061)
High SES	0.723***	0.519***	0.504***
	(0.016)	(0.015)	(0.015)
Repeated ×	0.100	0.197*	0.335***
High SES	(0.12)	(0.11)	(0.12)
6th grade		0.283***	0.264***
aspirations in years		(0.0044)	(0.0053)
Repeated ×			-0.107***
6th grade aspirations in years			(0.028)
High SES ×			0.0554***
6th grade aspirations in years			(0.0078)
Repeated ×			0.0694
High SES × 6th grade aspirations in years			(0.060)
Observations	61465	61465	61465
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 10: 8^{th} grade aspirations for those only passing mathematics at 6^{th} grade midterm

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.702***	-0.582***	-1.005***
	(0.089)	(0.088)	(0.16)
High SES	0.703***	0.452***	0.459***
	(0.0079)	(0.0070)	(0.0072)
Repeated ×	0.259	0.186	0.644***
High SES	(0.16)	(0.15)	(0.21)
6th grade		0.345***	0.337***
aspirations in years		(0.0021)	(0.0031)
Repeated ×			-0.164***
6th grade aspirations in years			(0.051)
High SES ×			0.0136***
6th grade aspirations in years			(0.0035)
Repeated ×			0.212***
High SES × 6th grade aspirations in years			(0.079)
Observations	309209	309209	309209
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 11: 8^{th} grade aspirations for those with higher mathematics marks at 6^{th} grade midterm

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0846*	-0.0847	-0.108**
	(0.051)	(0.052)	(0.049)
High SES	0.182***	0.150***	0.141***
	(0.029)	(0.029)	(0.030)
Repeated ×	0.0208	0.0317	0.0588
High SES	(0.11)	(0.11)	(0.11)
6th grade		0.0414***	0.0355***
aspirations in years		(0.0070)	(0.0087)
Repeated ×			-0.0596***
6th grade aspirations in years			(0.022)
High SES ×			0.0199
6th grade aspirations in years			(0.014)
Repeated ×			0.0861
High SES × 6th grade aspirations in years			(0.053)
Observations	3338	3338	3338
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 12: Attending a secondary school ending with a high school diploma (1 in mathematics)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0637**	-0.0519**	-0.0610*
	(0.026)	(0.025)	(0.033)
High SES	0.172***	0.136***	0.136***
	(0.0056)	(0.0055)	(0.0056)
Repeated ×	0.0451	0.0444	0.0516
High SES	(0.046)	(0.045)	(0.051)
6th grade		0.0531***	0.0536***
aspirations in years		(0.0015)	(0.0019)
Repeated ×			-0.00978
6th grade aspirations in years			(0.016)
High SES ×			-0.00115
6th grade aspirations in years			(0.0027)
Repeated ×			0.0163
High SES × 6th grade aspirations in years			(0.026)
Observations	37074	37074	37074
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 13: Attending a secondary school ending with a high school diploma (2 in mathematics)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.204***	-0.198***	-0.241***
	(0.047)	(0.047)	(0.072)
High SES	0.0841***	0.0654***	0.0440***
	(0.0018)	(0.0017)	(0.0016)
Repeated ×	0.148**	0.144**	0.194**
High SES	(0.061)	(0.060)	(0.080)
6th grade		0.0256***	0.0506***
aspirations in years		(0.00050)	(0.00087)
Repeated ×			-0.0293
6th grade aspirations in years			(0.023)
High SES ×			-0.0404***
6th grade aspirations in years			(0.00096)
Repeated ×			0.0774***
High SES × 6th grade aspirations in years			(0.029)
Observations	212904	212904	212904
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 14: Attending a secondary school ending with a high school diploma (3-5 in mathematics)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing vwill lalues. 6th grade aspirations are demeaned.

References

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 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.

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.

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.

Gary-Bobo, Robert J., Marion Goussé, and Jean-Marc Robin, "Grade retention and unobserved heterogeneity," *Quantitative Economics*, 2016, *7* (3), 781–820.

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.

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.

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).

Jacob, Brian A. and Lars Lefgren, "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *The Review of Economics and Statistics*, February 2004, *86* (1), 226–244.

_ **and** _ , "The Effect of Grade Retention on High School Completion," *American Economic Journal: Applied Economics*, 2009, *1* (3), 33–58.

Müller, Maximilian, "Intergenerational Transmission of Education: Internalized Aspirations versus Parent Pressure," Working Paper 2021.

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.

Pekkarinen, Tuomas, "School tracking and intergenerational social mobility," *IZA World of Labor*, 2018, (56).

Schütz, Gabriela, Heinrich W. Ursprung, and Ludger Wössmann, "Education policy and equality of opportunity," *Kyklos*, 2008, *61* (2), 279–308.

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

A Appendix

A.1 Descriptive tables

	Socio-economic status		
6th grade aspirations in 3 categories	Low SES	High SES	Total
At most vocational	0.0298	0.0213	0.0280
	40175	11056	51231
High school diploma or post-secondary non-tertiary	0.0104	0.0063	0.0084
	82594	80184	162778
College or university	0.0036	0.0014	0.0018
	41555	203903	245458
Total	0.0135	0.0035	0.0071
	164324	295143	459467

Table 15: Observations and share of repeaters in each aspiration-SES cell

Note: Each cell contains the share of repeaters in the cell and the number of observations in the cell. The sample for this table contains children who are in primary school in 6^{th} grade and who did not repeat the 6^{th} grade.

	Socio-economic status		
Repeated 7th grade	Low	High	Total
No	0.155	0.135	0.142
Yes	-0.154	0.070	-0.083
Total	0.151	0.135	0.141

Table 16: Change in the educational aspirations in years by parental education

Note: Each cell shows the average change in the aspired years of education between 6^{th} grade and 8^{th} grade in the cell.

A.2 Retention probabilities

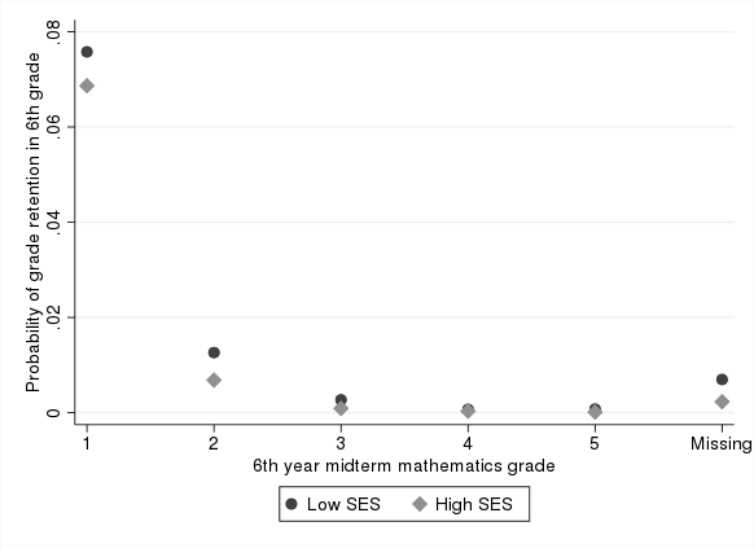

Figure 5: Probability of repeating 6th grade by 6th year midterm mathematics grade

	1: Low-SES non-rep	1: Low-SES rep	1: High-SES non-rep	1: High-SES rep	2: Low-SES non-rep	2:Low-SES rep	2: High-SES non-rep	2:High-SES rep
6th grade mathematics test	1294.82	1308.36	1357.22	1315.58	1341.54	1309.53	1386.18	1373.29
score	(150.55)	(142.96)	(149.16)	(113.89)	(150.24)	(139.27)	(143.01)	(164.63)
6th grade reading test score	1283.34	1272.77	1359.71	1282.60	1332.81	1286.81	1392.26	1345.19
	(153.47)	(135.00)	(165.11)	(158.66)	(152.78)	(141.26)	(152.78)	(147.82)
6th year midterm literature	2.40	2.12	2.63	2.18	2.90	2.43	3.17	2.59
grade	(0.80)	(0.73)	(0.94)	(0.93)	(0.80)	(0.69)	(0.81)	(0.70)
6th year midterm Hungarian	2.23	1.99	2.44	2.13	2.71	2.32	2.91	2.56
grammar grade	(0.76)	(0.65)	(0.83)	(0.81)	(0.75)	(0.66)	(0.77)	(0.71)
Parents lived together in 6th	0.72	0.76	0.65	0.50	0.75	0.64	0.70	0.45
grade	(0.45)	(0.43)	(0.48)	(0.51)	(0.43)	(0.48)	(0.46)	(0.50)
Parents separated from 6th to	0.06	0.12	0.06	0.18	0.05	0.10	0.06	0.10
8th grade	(0.24)	(0.32)	(0.23)	(0.39)	(0.23)	(0.30)	(0.24)	(0.31)
Mother does not work in 6th	0.51	0.51	0.25	0.21	0.44	0.46	0.20	0.24
grade	(0.50)	(0.50)	(0.43)	(0.41)	(0.50)	(0.50)	(0.40)	(0.43)
Mother has a permanent job in	0.35	0.36	0.64	0.66	0.43	0.41	0.68	0.62
6th grade	(0.48)	(0.48)	(0.48)	(0.48)	(0.49)	(0.49)	(0.47)	(0.49)
Father does not work in 6th	0.24	0.23	0.10	0.18	0.19	0.24	0.09	0.11
grade	(0.43)	(0.43)	(0.30)	(0.39)	(0.39)	(0.43)	(0.29)	(0.32)
Father has a permanent job in	0.53	0.57	0.69	0.63	0.58	0.53	0.71	0.69
6th grade	(0.50)	(0.50)	(0.46)	(0.49)	(0.49)	(0.50)	(0.46)	(0.47)
Mother stopped working between	0.08	0.06	0.07	0.13	0.08	0.09	0.05	0.06
6th and 8th grades	(0.27)	(0.25)	(0.25)	(0.34)	(0.27)	(0.29)	(0.22)	(0.24)
Father stopped working between	0.07	0.09	0.04	0.05	0.07	0.06	0.04	0.03
6th and 8th grades	(0.26)	(0.28)	(0.21)	(0.23)	(0.26)	(0.25)	(0.20)	(0.17)
Number of GP visits in 7th	5.78	7.06	5.07	4.16	5.42	7.75	4.60	7.28
grade	(5.15)	(6.20)	(4.89)	(3.08)	(5.10)	(6.74)	(4.38)	(7.29)
Days spent in hospital in 7th	0.35	0.41	0.46	0.24	0.30	0.65	0.32	1.50
grade	(2.72)	(1.36)	(2.19)	(0.85)	(1.97)	(4.65)	(1.79)	(8.90)
Observations	1418	94	592	38	13720	296	9090	96

Table 17: Characteristics of repeaters by 6^{th} year mathematics grades and SES: failing (1) and just passing (2) mathematics

	3-5: Low-SES non-rep	3-5: Low-SES rep	3-5: High-SES non-rep	3-5: High-SES rep
6th grade mathematics test	1481.94	1353.92	1580.84	1477.61
score	(164.12)	(163.08)	(166.98)	(162.14)
6th grade reading test score	1469.79	1345.19	1578.84	1464.13
	(165.77)	(169.70)	(166.89)	(187.27)
6th year midterm literature	3.92	2.98	4.39	3.52
grade	(0.87)	(0.94)	(0.73)	(1.07)
6th year midterm Hungarian	3.72	2.95	4.17	3.49
grammar grade	(0.87)	(0.86)	(0.80)	(1.05)
Parents lived together in 6th	0.78	0.69	0.79	0.55
grade	(0.41)	(0.47)	(0.40)	(0.50)
Parents separated from 6th to	0.05	0.13	0.04	0.12
8th grade	(0.22)	(0.33)	(0.20)	(0.33)
Mother does not work in 6th	0.34	0.49	0.14	0.23
grade	(0.47)	(0.50)	(0.35)	(0.42)
Mother has a permanent job in	0.52	0.43	0.75	0.61
6th grade	(0.50)	(0.50)	(0.44)	(0.49)
Father does not work in 6th	0.14	0.24	0.06	0.11
grade	(0.34)	(0.43)	(0.24)	(0.31)
Father has a permanent job in	0.66	0.53	0.72	0.62
6th grade	(0.47)	(0.50)	(0.45)	(0.49)
Mother stopped working between	0.07	0.10	0.04	0.10
6th and 8th grades	(0.25)	(0.30)	(0.19)	(0.30)
Father stopped working between	0.05	0.10	0.03	0.06
6th and 8th grades	(0.22)	(0.30)	(0.16)	(0.24)
Number of GP visits in 7th	4.81	7.51	3.99	6.16
grade	(4.69)	(7.23)	(4.03)	(7.43)
Days spent in hospital in 7th	0.27	1.36	0.24	2.20
grade	(1.64)	(6.89)	(1.69)	(11.02)
Observations	39674	103	102489	82

Table 18: Characteristics of repeaters by 6^{th} year mathematics grades and SES: grades 3-5

A.3 Robustness checks for main regressions

A.3.1 Using three categories of SES

	Aspirations in 8th grade: at most vocational school				
	(1)	(2)	(3)	(4)	(5)
Repeated	0.387***	0.378***	0.255***	0.153***	0.156***
	(0.011)	(0.013)	(0.013)	(0.012)	(0.012)
Medium SES		-0.179***	-0.104***	-0.0823***	-0.0688***
		(0.0024)	(0.0017)	(0.0014)	(0.0013)
High SES		-0.209***	-0.111***	-0.0819***	-0.0632***
		(0.0025)	(0.0016)	(0.0013)	(0.0013)
Repeated × Medium SES		-0.123***	-0.109***	-0.113***	-0.111***
		(0.025)	(0.023)	(0.022)	(0.022)
Repeated × High SES		-0.247***	-0.193***	-0.179***	-0.183***
		(0.026)	(0.023)	(0.023)	(0.023)
6th grade aspirations, baseline: at most vocational					
High school diploma or			-0.376***	-0.289***	-0.275***
post-secondary non-tertiary			(0.0031)	(0.0030)	(0.0029)
College or university			-0.430***	-0.294***	-0.273***
			(0.0034)	(0.0031)	(0.0030)
Constant	0.0991***	0.223***	0.515***	0.878***	0.813***
	(0.0019)	(0.0026)	(0.0036)	(0.032)	(0.032)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

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

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

Table 19: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma. Medium SES means at least one parent has a high school diploma. High SES means at least one parent has a tertiary degree. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	Aspirations in 8th grade: tertiary				
	(1)	(2)	(3)	(4)	(5)
Repeated	-0.428***	-0.215***	-0.0768***	0.000326	-0.000685
	(0.0072)	(0.0053)	(0.0057)	(0.0055)	(0.0056)
Medium SES		0.318***	0.167***	0.131***	0.115***
		(0.0025)	(0.0021)	(0.0019)	(0.0018)
High SES		0.581***	0.316***	0.251***	0.212***
		(0.0029)	(0.0027)	(0.0025)	(0.0021)
Repeated × Medium SES		-0.204***	-0.131***	-0.0931***	-0.0917***
		(0.017)	(0.017)	(0.017)	(0.017)
Repeated × High SES		-0.156***	-0.110***	-0.0564**	-0.0468*
		(0.032)	(0.027)	(0.025)	(0.024)
6th grade aspirations, baseline: at most vocational					
High school diploma or			0.146***	0.0496***	0.0412***
post-secondary non-tertiary			(0.0019)	(0.0018)	(0.0019)
College or university			0.620***	0.387***	0.365***
			(0.0023)	(0.0026)	(0.0026)
Constant	0.541***	0.254***	0.0178***	-0.505***	-0.417***
	(0.0043)	(0.0021)	(0.0019)	(0.025)	(0.026)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

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

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

Table 20: Probability of 8^{th} grade aspirations being tertiary education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma. Medium SES means at least one parent has a high school diploma. High SES means at least one parent has a tertiary degree. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

A.3.2 Using continuous SES

	Aspirations in 8th grade: at most vocational school				
	(1)	(2)	(3)	(4)	(5)
Repeated	0.387***	0.278***	0.169***	0.0702***	0.0732***
	(0.011)	(0.011)	(0.0097)	(0.0096)	(0.0094)
Maximum years of education of		-0.0358***	-0.0195***	-0.0141***	-0.0111***
parents		(0.00056)	(0.00038)	(0.00029)	(0.00028)
Repeated × Maximum		-0.0376***	-0.0299***	-0.0279***	-0.0275***
years of education of parents		(0.0037)	(0.0035)	(0.0034)	(0.0034)
6th grade aspirations, baseline: at most vocational					
High school diploma or			-0.384***	-0.296***	-0.280***
post-secondary non-tertiary			(0.0031)	(0.0030)	(0.0029)
College or university			-0.437***	-0.301***	-0.279***
			(0.0033)	(0.0031)	(0.0030)
Constant	0.0991***	0.100***	0.453***	0.822***	0.772***
	(0.0019)	(0.0012)	(0.0033)	(0.032)	(0.032)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

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

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

Table 21: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. SES is proxied with the years of education of the highest educated parent. Parental education is demeaned. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	Aspirations in 8th grade: tertiary				
	(1)	(2)	(3)	(4)	(5)
Repeated	-0.428***	-0.340***	-0.154***	-0.0494***	-0.0454***
	(0.0072)	(0.010)	(0.0094)	(0.0088)	(0.0087)
Maximum years of education of		0.0948***	0.0514***	0.0398***	0.0331***
parents		(0.00032)	(0.00034)	(0.00033)	(0.00031)
Repeated × Maximum		-0.0371***	-0.0244***	-0.0143***	-0.0126***
years of education of parents		(0.0039)	(0.0033)	(0.0031)	(0.0030)
6th grade aspirations, baseline: at most vocational					
High school diploma or			0.148***	0.0537***	0.0451***
post-secondary non-tertiary			(0.0021)	(0.0019)	(0.0019)
College or university			0.630***	0.399***	0.377***
			(0.0024)	(0.0026)	(0.0026)
Constant	0.541***	0.537***	0.164***	-0.373***	-0.313***
	(0.0043)	(0.0018)	(0.0022)	(0.025)	(0.026)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

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

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

Table 22: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. SES is proxied with the years of education of the highest educated parent. Parental education is demeaned. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	8th grade educational aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-2.462***	-1.904***	-0.983***	-1.343***	-0.528***	-0.531***
	(0.043)	(0.052)	(0.045)	(0.057)	(0.054)	(0.053)
Maximum years of		0.415***	0.194***	0.196***	0.140***	0.115***
education of parents		(0.0022)	(0.0017)	(0.0017)	(0.0015)	(0.0014)
Repeated ×		-0.0509***	0.00187	0.132***	0.112***	0.116***
Maximum years of education of parents		(0.019)	(0.017)	(0.022)	(0.021)	(0.021)
6th grade			0.528***	0.525***	0.341***	0.319***
aspirations in years			(0.0019)	(0.0019)	(0.0019)	(0.0019)
Repeated ×				-0.162***	-0.0862***	-0.0844***
6th grade aspirations in years				(0.021)	(0.020)	(0.019)
Maximum years of				-0.00544***	-0.00379***	-0.00437***
education of parents × 6th grade aspirations in years				(0.00069)	(0.00063)	(0.00062)
Repeated ×				0.0468***	0.0294***	0.0301***
Maximum years of education of parents × 6th grade aspirations in years				(0.0071)	(0.0067)	(0.0067)
Constant	14.71***	14.70***	14.65***	14.66***	11.57***	11.69***
	(0.019)	(0.0081)	(0.0046)	(0.0056)	(0.15)	(0.15)
Observations	391412	391412	391412	391412	391412	391412
6th grade controls	no	no	no	no	yes	yes
Year fixed effect	no	no	no	no	yes	yes
School fixed effect	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 23: Educational aspirations in 8^{th} grade in years

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. SES is proxied with the years of education of the highest educated parent. Both the 6th grade aspirations variable and the parental education variable are demeaned. In column 4, 6th grade aspirations are added as controls. In column 5, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 6 adds 6th grade school fixed effects.

A.3.3 Regression of continuous aspirations on the Admin3 data

	8th grade educational aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated 7th grade	-2.826***	-1.938***	-1.034***	-1.751***	-0.790***	-0.801***
	(0.078)	(0.071)	(0.068)	(0.16)	(0.15)	(0.15)
High SES		2.255***	1.054***	1.092***	0.643***	0.501***
		(0.023)	(0.015)	(0.017)	(0.014)	(0.014)
Repeated 7th grade		-0.327*	-0.00860	0.680***	0.673***	0.617**
× High SES		(0.19)	(0.16)	(0.25)	(0.25)	(0.24)
6th grade			0.536***	0.500***	0.331***	0.316***
aspirations in years			(0.0027)	(0.0047)	(0.0046)	(0.0046)
Repeated 7th grade				-0.210***	-0.123***	-0.121***
× 6th grade aspirations in years				(0.044)	(0.042)	(0.042)
High SES ×				0.0543***	0.0718***	0.0644***
6th grade aspirations in years				(0.0054)	(0.0051)	(0.0050)
Repeated 7th grade				0.163*	0.0690	0.0682
× High SES × 6th grade aspirations in years				(0.085)	(0.079)	(0.078)
Constant	15.27***	13.80***	14.57***	14.52***	10.61***	10.90***
	(0.024)	(0.017)	(0.012)	(0.015)	(0.29)	(0.29)
Observations	194852	194852	194852	194852	194852	194852
6th 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 class level. Clustered standard errors in parentheses.

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

Table 24: Educational aspirations in 8^{th} grade in years, Admin3 data

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. The 6th grade aspirations variable is demeaned. In column 4, 6th grade aspirations are added as controls. In column 5, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 6 adds 6th grade school fixed effects.

A.3.4 Regressions on subsamples with health controls

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.180	-0.139	-0.239
	(0.31)	(0.31)	(0.34)
High SES	0.849***	0.631***	0.572***
	(0.15)	(0.15)	(0.16)
Repeated ×	-0.258	-0.313	-0.0608
High SES	(0.59)	(0.57)	(0.64)
6th grade		0.280***	0.244***
aspirations in years		(0.048)	(0.055)
Repeated ×			-0.122
6th grade aspirations in years			(0.17)
High SES ×			0.124
6th grade aspirations in years			(0.091)
Repeated ×			-0.0762
High SES × 6th grade aspirations in years			(0.37)
Observations	2772	2772	2772
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 25: 8^{th} grade aspirations for those failing mathematics at 6^{th} grade midterm, Admin3 database

Note: Controls include gender, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.512***	-0.459***	-0.547***
	(0.083)	(0.079)	(0.088)
High SES	0.746***	0.555***	0.540***
	(0.027)	(0.026)	(0.026)
Repeated ×	0.354	0.405*	0.522**
High SES	(0.22)	(0.21)	(0.21)
6th grade		0.260***	0.236***
aspirations in years		(0.0083)	(0.010)
Repeated ×			-0.0818**
6th grade aspirations in years			(0.036)
High SES ×			0.0601***
6th grade aspirations in years			(0.015)
Repeated ×			0.0473
High SES × 6th grade aspirations in years			(0.13)
Observations	28193	28193	28193
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 26: 8^{th} grade aspirations for those just passing mathematics at 6^{th} grade midterm, Admin3 database

Note: Controls include gender, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.522***	-0.309*	-0.566
	(0.17)	(0.16)	(0.36)
High SES	0.769***	0.476***	0.523***
	(0.015)	(0.014)	(0.016)
Repeated ×	0.0109	-0.143	0.165
High SES	(0.33)	(0.30)	(0.46)
6th grade		0.378***	0.337***
aspirations in years		(0.0030)	(0.0054)
Repeated ×			-0.0605
6th grade aspirations in years			(0.089)
High SES ×			0.0581***
6th grade aspirations in years			(0.0059)
Repeated ×			0.0883
High SES × 6th grade aspirations in years			(0.13)
Observations	157406	157406	157406
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 27: 8^{th} grade aspirations for those with higher mathematics marks at 6^{th} grade midterm, Admin3 database

Note: Controls include gender, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0986	-0.0799	-0.189*
	(0.13)	(0.12)	(0.11)
High SES	0.220***	0.195***	0.204***
	(0.064)	(0.064)	(0.065)
Repeated ×	-0.169	-0.228	-0.202
High SES	(0.28)	(0.29)	(0.31)
6th grade		0.0415***	0.0515***
aspirations in years		(0.015)	(0.019)
Repeated ×			-0.143***
6th grade aspirations in years			(0.050)
High SES ×			-0.0216
6th grade aspirations in years			(0.029)
Repeated ×			0.238
High SES × 6th grade aspirations in years			(0.18)
Observations	1434	1434	1434
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 28: Attending a secondary school ending with a high school diploma (1 in mathematics, Admin3 database)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0761*	-0.0674*	-0.0717
	(0.042)	(0.041)	(0.052)
High SES	0.158***	0.132***	0.134***
	(0.0084)	(0.0084)	(0.0085)
Repeated ×	-0.0139	-0.0179	-0.0195
High SES	(0.070)	(0.068)	(0.075)
6th grade		0.0385***	0.0420***
aspirations in years		(0.0020)	(0.0029)
Repeated ×			-0.00619
6th grade aspirations in years			(0.026)
High SES ×			-0.00725**
6th grade aspirations in years			(0.0037)
Repeated ×			0.0148
High SES × 6th grade aspirations in years			(0.030)
Observations	17540	17540	17540
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 29: Attending a secondary school ending with a high school diploma (2 in mathematics, Admin3 database)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.224***	-0.220***	-0.107
	(0.073)	(0.072)	(0.072)
High SES	0.0745***	0.0678***	0.0492***
	(0.0023)	(0.0023)	(0.0021)
Repeated ×	0.116	0.110	-0.00263
High SES	(0.093)	(0.092)	(0.090)
6th grade		0.00865***	0.0264***
aspirations in years		(0.00030)	(0.00074)
Repeated ×			0.0279**
6th grade aspirations in years			(0.014)
High SES ×			-0.0247***
6th grade aspirations in years			(0.00078)
Repeated ×			-0.00121
High SES × 6th grade aspirations in years			(0.021)
Observations	114298	114298	114298
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	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 30: Attending a secondary school ending with a high school diploma (3-5 in mathematics, Admin3 database)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

A.4 Data Appendix

The student level questionnaires and test score data are available between 2008 and 2017, with in total around 2.8 million observations. I am interested in those students for whom at least 6^{th} and 8^{th} grade data can be linked, so as a first step, I drop all students whose data cannot be matched between these two grades. Then, I create an indicator variable for whether the student repeated any grade up to the 6^{th} grade. I use two variables here: whether the student repeated in the lower grades of primary school, and whether she repeated in the higher grades of primary school, both reported in grade 6. I classify the student as no repeater if both variables equal 1, which means no repetition in those particular grades. Unfortunately, in 6^{th} grade, 20 percent of this constructed measure are missing. I recover this data from the 8^{th} grade and 10^{th} grade surveys, wherever possible, since later surveys also include these questions. When reporting about repetition in lower grades was inconsistent through survey years, I use the mode of the answers to impute the 6^{th} grade value. In the end, I have self-reported information on repetition up to $6th$ grade from 94.3 percent of the 6^{th} grade sample. Because of the compulsory school starting age rules in Hungary, students are 12 or 13 years old in the year they finish 6^{th} grade (that is the time of the test), but in a small percentage of cases it is possible that students – without repeating a grade – turn 14 in this year. Therefore, I drop students younger than 12 when writing the 6^{th} grade test, and assign students older than than 14 to the repeater group.[9] This is around 1 percent of the sample. I construct a wide database where the time variable is the grade, and I keep the first observations from 6^{th}, 8^{th} and 10^{th} grades. I define a variable for each grade indicating if the student had multiple observations from that particular grade, meaning that she repeated it, and another variable that counts how many observations the student has from the grade.

A.4.1 Birth year and month and school starting age variables

The source of birth year and month can be regarded as administrative data, but it contains missing values across years (less than 1 percent of all observations). I recover missing data from the other years' data of the same person. School starting age is a constructed variable from the self-reported year when students started school and their birth year and month. School starting age means the age the child had already turned when started school. Datasets before 2015 only contain this constructed school starting age, while from 2015, they also include the school starting year as students reported it. For both variables, I first create a corrected variable that adds the last non-

[9] I use the first occurrence in the 6^{th} grade sample, since a fraction of students had to repeat 6^{th} grade, which resulted in them appearing multiple times in the 6^{th} grade samples.

missing value to all observations of the same student. Then, as later I would only like to use the school starting year, I calculate school starting year from the corrected school starting age variable for students for whom the year was missing. As school starts in September, I regard students up to September as having turned the age they turn in that year when they started school, while students born between October and December, their school starting age is a year younger than the age they turn in the particular year.

A.4.2 Creating years of education from the categorical education variables

I create continuous variables from two categorical education variables: parental education and educational aspirations. For both variables I use the following coding: unfinished primary education = 7 years, primary education = 8 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 similarly as 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 a large overestimation of their educational preferences.

A.4.3 Admin3 NABC extension

I took the same data cleaning steps as for the National ABC until keeping the last observations and the reshape wide. I use all observations for the admin data and keep the data in a long format to be able to merge + append with the admin panel. I first matched the observations from the National ABC extension to data from May in the same year in the administrative database. I 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 who were older than 22 (the highest age in the National ABC) in the matched database according to the administrative database. I checked if the rest of the observations are matched well in terms of age: I considered a match bad if KOR from the Admin3 database were higher than AGEATTEST + 1 or lower than AGEATTEST - 1 from the National ABC database. First I corrected the birth year for those observations where the age mismatch was the result of inconsistent reporting of the birth year. At the end I still had 40 observations with age mismatch. I dropped these 40 observations, too. I did the same steps as with the National ABC database, but here some data were missing: there was no data on class and on special education needs and valid/ not valid test status. I merged the National ABC data to the monthly main 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., number of visits to the GP in the 6^{th} grade, 7^{th} grade, etc. I kept the first observations of 6^{th} grade, 8^{th}, and 10^{th} grade NABC data and the corresponding data from Admin3 from these years. I also kept the 7^{th} grade healthcare data for predicting retention.