If at First You Don't Succeed: A Dynamic Evaluation of Grade Retention

Emilio Borghesan, Hugo Reis, and Petra E. Todd [1]

February 19, 2025

[1]Emilio Borghesan is an Assistant Professor in the Economics Department at the University of Michigan. Hugo Reis is a research economist at the Banco de Portugal and a Professor of Economics at Catolica Lisbon SBE and a member of IZA. Petra Todd is Edmund J. and Louise W. Kahn Term Professor of Economics at the University of Pennsylvania and a member of the U Penn Population Studies Group, NBER, HCEO and IZA. This paper was presented in seminars at U Penn and Brandeis University, SOLE 2022 Conference, and as a keynote lecture at the Brazilian Econometric Society meetings and the Barcelona School of Economics Summer Forum. We thank participants for many helpful comments. Hugo Reis acknowledges the financial support of Fundação para a Ciência e Tecnologia through the project number PTDC/EGE-ECO/6055/2020.

Abstract

In many European countries, it is common for secondary school students to be retained. This paper evaluates the impact of grade retention on learning, dropout, and educational attainment in Portugal, which has some of the highest retention rates in the world. An extended Roy model is developed and implemented to model retention's cumulative effects on test scores, dropout, and educational attainment for different subgroups followed over multiple grades. Results show substantial heterogeneity in test score impacts. In math, 82% of students retained experience test score gains, with an average impact of 0.46 sd among those who graduate; in Portuguese, only 42% of students experience gains, with an average impact of -0.17 sd. Also, retention increases dropout by 7 pp for the average retained student. We validate our model by simulating retention effects on dropout for 12th grade students at the margin of being retained for whom we can compare model-based estimates to RDD estimates. Lastly, the estimated model is used to solve for the optimal retention policy that maximizes average lifetime earnings, considering that retention yields cognitive gains for some students who stay in school while also decreasing educational attainment for students who drop out. Portuguese retention rates are found to be too high for plausible estimates of cognitive skill returns.

1 Introduction

Grade retention patterns differ markedly across countries. Although retention is rare in the Nordic countries, South Korea, Japan, and Malaysia, it is more common in the Netherlands, where 20.1% of students repeat a grade by age 15, France (22.1%), Portugal (31.0%), and Belgium (34.0%). In Portugal, the setting for this paper, retention is especially prevalent in secondary school, where over 40% of students who commenced tenth grade in 2009/2010 were made to repeat a grade. Grade retention is expensive, because retained students must be educated for another year, and schools only avoid these costs if retained students drop out. Whether retention has benefits that outweigh the costs is a question of interest to both policymakers and educators.

Ethical concerns make it infeasible to evaluate grade retention using randomized controlled trials, so prior studies are entirely based on observational data, which pose two distinct methodological challenges. The first is that retained students are typically negatively selected on both observable and unobservable dimensions, such as ability, socioeconomic background, intrinsic motivation, and emotional maturity. The second challenge is that retention may increase the likelihood of dropping out, particularly in later grades, which may cause dynamic sample selection bias in comparisons of retained and promoted students. The goal of this study is to analyze effects of secondary school grade retention using a framework that controls for the endogeneity of the retention decision and for nonrandom dropout.

We address these challenges by developing and estimating a structural model to analyze retention effects on academic achievement and dropout. Our model builds on the generalized two-sector Roy model where the sectors correspond to being retained or not (Roy, 1951; Heckman and Honore, 1990; Heckman and Vytlacil, 2007). We extend the model by including lagged test scores to capture the dynamic accumulation of knowledge across grades in response to retention and each student's history of educational inputs, and by including a second selection equation to control for potential sample selection bias caused by dropout. Our approach sheds light on the mechanisms through which retention affects test scores. For example, retained students who remain in school take one year longer to graduate, and our analysis distinguishes whether the causal effect of retention on test scores stems from the additional year of educational inputs, the differential productivity of educational inputs in the year the student is retained, or an extra year of skill depreciation.

Much of the grade retention literature is based on primary and middle school grades where school dropout is not a key concern. In contrast, our model framework accounts for sample selection arising from students dropping out of school, which is particularly important in secondary school grades. 34% percent of all students, and fully 77% of retained students, who enroll in Portuguese secondary schools drop out. Our model provides a unified framework to jointly analyze retention effects on educational attainment and academic achievement. It incorporates rich observed heterogeneity by allowing outcomes to depend on student and family demographics as well as on observable dimensions of school and teacher quality. The model also allows the unobservables that jointly enter the multiple outcome equations (test scores, dropout, and retention) to be correlated.[1]

Our model is estimated using a large administrative database (MISI) from the Ministry of Education in Portugal that contains information on annual school enrollment and standardized test scores in math and Portuguese for approximately fifty thousand students for the years 2008-2013. Enrollment and grade retention are measured in every year, but the standardized tests are high-stakes tests that are only administered in grades 9 and 12. If a student is retained in 9th or 12th grade, we observe multiple test scores corresponding to the different years when the student took the same-grade test. Our model and estimation approach accommodates the fact that some students have multiple test scores in the same grade as well as the fact that enrollment, dropout, and retention are observed annually whereas test scores are observed only in the grades when students take the tests. Educational researchers commonly encounter such data complications, because standardized tests are often not taken in every grade.

[1]As further described below, we invoke exclusion restrictions to obtain identification of the unobservable covariances.

Our estimation strategy addresses this differential timing in a way that is consistent with a theoretical knowledge accumulation model where the productivity of inputs may vary depending on whether the student has been promoted and is seeing the material for the first time or is repeating the grade. We derive moments from the model that correspond to covariances between dependent variables (test scores, retention, and dropout) and variables that our model takes as exogenous for subgroups of students with different schooling trajectories. These moments form the basis for efficient parameter estimation via simulated method of moments (SMM).

We use the estimated model to evaluate the mean and distributional effects of grade retention in Portuguese secondary schools. We find that the status quo retention policy produces, on average, a gain of 0.46 sd in 12th grade math scores and a loss of 0.17 sd in Portuguese scores for retained students who ultimately graduate within four years. We find substantial heterogeneity in retention effects across students, with 82% (42%) of students retained in math (Portuguese) experiencing positive test score effects. We find evidence of positive sorting into retention: average test score impacts for retained students (treatment-on-the treated, or TT) are larger than average treatment effects (ATEs) corresponding to retention in each grade. Sorting is based on both observable and unobservable dimensions of heterogeneity. For example, students who enter secondary school with ninth grade math scores in the bottom tercile of the distribution are both more likely to be retained and to have higher causal effects of retention than students with higher initial test scores. There is also sorting on unobserved shocks to math test scores. The upshot of positive sorting on gains is that, when retention rates are low, retained students will be more likely to benefit from retention.

A possible concern in using our estimated model to evaluate a range of alternative retention policies is model mispecification. To address this concern and investigate model fit, we compare the test score impacts simulated from the model to regression discontinuity design (RDD) estimates obtained directly from the data for a subgroup of 12th graders at the margin of being retained. The RDD estimates exploit the fact that retention in the 12th grade is subject-specific and depends, in part, on a weighted average of course grades and scores on nationwide tests. Using the discontinuities in subject-specific retention probabilities, we obtain fuzzy RDD estimates of the 12th grade retention effect in each subject on the probability of dropping out in that subject. The RDD estimates are a statistically insignificant 12.1 percentage point (pp) increase in math and a significant 52.5 pp increase in Portuguese in the data. According to model simulations, 12th grade retention in math (Portuguese) increases the probability of dropout by 11.6 (42.7) pp at the margin. The estimates generated by the model are within the confidence intervals of the estimated RDDs from the data.

A well-known drawback of RDD estimators is that they only identify causal effects for individuals near the margin, in this case, for students at the margin of being retained. When we use the model to estimate the average effect of treatment on the treated (TT) and average treatment effect (ATE), we find that retention effects are lower for students away from the margin. The RDD therefore overestimates both TT and ATE. Selection plays a strong role: The correlation coefficient between shocks to retention and shocks to dropout is 0.635, meaning that many retained students would have dropped out even if they had not been retained. A second mechanism that blunts retention's deleterious effects on dropout away from the margin is positive sorting on test score gains into retention. Students who are more likely to be retained have higher test score gains, which reduces the drop out probability.

In the last section of the paper, we perform a cost-benefit assessment of whether the lifetime earnings benefits of grade retention exceed the costs, using the estimated heterogeneous impacts of retention on test scores and dropout. Retained students who stay in school begin full-time work later than promoted students. Our cost-benefit analysis trades off the benefits of increased math test scores with the costs of reduced educational attainment (for those who drop out) and delayed labor market entry (for those who complete secondary school). To carry out this analysis, we link the MISI data with Portugal's employer-employee matched data set, the *Quadros de Pessoal* (QP), at the concelho (municipality) level.[2] We regress

[2]The *Quadros de Pessoal* data are used frequently in the analysis of the Portuguese labor market. See, for example, Card et al. (2016) on how sorting to firms affects the gender pay gap and Portugal et al. (2024) on the analysis of the returns to schooling.

average wages in each concelho on average test scores in math and Portuguese and allow different wage-experience profiles for secondary school dropouts and graduates. Given our estimates of the returns associated with increased test scores, we find that it is not optimal to retain any students. The cost of dropout and delayed labor market entry outweigh the wage returns to greater skills. We then ask, what is the minimal return to math skill necessary to justify retaining a positive fraction of students? This estimate, of nearly 30% per standard deviation (sd) in math test scores, is far higher than estimates typically found in the literature (Murnane et al., 2000; Cawley et al., 2001; Dougherty, 2003; Heckman et al., 2006; Chetty et al., 2011; Watts, 2020). Our cost-benefit analysis is conservative in that it considers the costs and benefits for students only and does not account for reeducation costs for retained students. Our analysis therefore suggests that it may be prudent for high-retention countries to reduce retention rates.

2 Related Literature

Proponents of grade retention argue that the practice provides students with the opportunity to master the curriculum before moving on to more advanced material. Under this view, academic achievement is a cumulative process and mastering the material in one grade facilitates learning in the next. Opponents express concerns that retained students may be stigmatized, have a hard time adjusting to a new peer group and suffer from reduced self-esteem. If students get discouraged then high retention rates could inhibit learning and increase dropout. A large literature analyzes retention effects on test scores and dropout using a variety of empirical approaches. Sometimes retained students are compared to non-retained students of the same age and, at other times, they are compared to students in the same grade but of different ages.

The literature provides mixed evidence on whether grade retention is harmful or beneficial. Holmes and Matthews (1984) and Jimerson (2001) present meta-analyses that focus on the frequency of positive and negative estimated effects across studies without accounting for differences in research designs. Both papers conclude that the preponderance of the evidence is that grade retention is harmful. Allen et al. (2009) carry out a meta-analysis of a smaller set of 22 studies with well-matched comparison groups. They explore how estimated retention effect sizes vary with the quality of the study design, with the grade in which the student is retained, and with the number of grades since retention. Their findings challenge the view that retention is harmful. Below we discuss more recent literature, organizing the studies according to methodological approaches.

Several studies use regression-discontinuity (RD) designs, exploiting discontinuities in the rules determining which children are retained. For example, Jacob and Lefgren (2004) and Jacob and Lefgren (2009) use an RD estimator that exploits an accountability policy that was introduced in Chicago Public Schools. They find a modest benefit of third grade retention on achievement scores but no effect of sixth grade retention. When Jacob and Lefgren (2009) look at longer-term impacts, they find that eighth grade retention decreases high school completion. Manacorda (2012) analyzes grade retention effects in junior high school using administrative data from Uruguay, also exploiting a discontinuity in retention rules. He also finds that retention leads to an increase in dropout. Eren et al. (2017) use an RD design to analyze the net effects of summer school remediation and test-based promotion policies in Louisiana on high school completion and juvenile crime. For eighth grade students, they find that retention decreases crime but increases dropout.[3] A well-known limitation of RD designs is that they identify retention impacts only for the subgroup of children at the margin of being retained. In this paper, we also obtain RD estimates, which we can

[3]Several studies use an RD approach to analyze retention in earlier grades. Schwerdt et al. (2017) exploit a discontinuity in Florida's test-based promotion policy and find that third grade retention has large positive effects on achievement through grade 10 but that the effects fade out if students are compared to same-age peers. They also find that third grade retention increases students' grade point averages and leads them to take fewer remedial courses in high school but does not affect students' graduation probability. Figlio and Özek (2020) use an RD design to study a combined third grade retention and instructional support policy for English language learners in 12 Florida school districts. They find that grade retention increases language proficiency and the likelihood that students take more advanced coursework in middle school and high school. Winters and Greene (2012) study a program in Florida that compelled retained students to attend summer school and then assigned them to a high-quality teacher during the retention year. The study finds a statistically significant positive impact on student achievement in math, reading, and science that is sustained for some years after the treatment but then dissipates. Zhong (2024) finds that retention in the third grade in Texas reduces earnings between the ages of 23 and 25 by 22%.

compare to our model-based estimates as a way of validating the model. After finding that the estimates are similar, we then use the model to estimate retention effects for the full population and for various subgroups.

Another branch of the literature evaluates effects of grade retention using instrumental variables (IV) estimators, usually within a regression framework. When impacts are heterogeneous across students, IV estimates are interpretable as a local average treatment effect (LATE), which is the average impact of retention for children who were retained because of the value of the instrument (the so-called complier group).[4] An early study by Eide and Showalter (2001) uses the High School and Beyond data set to analyze the effects of high school grade retention on dropout and on subsequent earnings. The instruments are derived from kindergarten school entry rules and correspond to various functions of the difference between the child's birthday and the cutoff for starting kindergarten. Their OLS estimates suggest that grade retention increases the probability of dropping out of school and negatively affects earnings, but their IV estimates tend not to be statistically significantly different from zero. Pereira and Reis (2014) and Garcia-Pérez et al. (2014) study the impact of grade retention in Portugal and Spain using the PISA data set, also using birth date as an instrumental variable. Pereira and Reis (2014) find that grade retention has small positive impacts on educational outcomes in the short term, while Garcia-Pérez et al. (2014) find negative effects.

Another group of studies develops and applies factor analytic dynamic models (FADM).[5] For example, Fruehwirth et al. (2016) analyze the effects of grade retention in kindergarten and other elementary grades using ECLS-K data. The authors develop a potential outcomes framework in which retention's effects can vary with the grade in which the student was retained as well as the number of years since retention.[6] They find significant negative test

[4]See Angrist and Imbens (1995).

[5]Early discussion of these methods include Carneiro et al. (2003) and Heckman and Navarro (2007).

[6]Along with the outcomes model, they specify a probabilistic grade retention equation, where a low-dimensional set of unobservable factors can affect both the outcome and retention equations. Their impact analysis compares retained and nonretained children at the same age, which is possible because the ECLS-K tests are comparable across years. As they note, with age held constant, retained children were exposed to less curricula.

score effects on retained children. Saltiel and Sarzosa (2020) also use the ECLS-K data, analyzing retention effects in kindergarten and first grade. They estimate a dynamic model of cognitive and noncognitive skill formation, which incorporates children's latent abilities, parental skills, and investment choices.[7] Their results show that retention lowers cognitive skills for retained students, slightly increases noncognitive skills (by 0.02 s.d.) and increases parental investments.[8] Gary-Bobo et al. (2016) use a factor analytic framework to study the effects of grade retention on French lower secondary school students, grades 6-9. They find negative average treatment effects of retention on both math and French test scores, but small positive estimates (< 0.1 sd) of the average effect of treatment on the treated for both subjects. Cockx et al. (2019) develop and estimate a FADM to estimate retention effects in Flemish secondary schools on test scores, dropout, downgrading of schooling track, and delayed graduation. They find neutral effects on short-term academic achievement but longer-term adverse effects on the other schooling outcomes, particularly for less able students. De Groote (2024) uses a dynamic discrete choice model with endogenous effort to analyze grade retention and forced downgrading, also in Flemish secondary schools. He finds that a policy promoting weaker students to the next grade in a less academic track reduces dropout relative to requiring them to repeat the grade in the same track.

Lastly, some studies exploit policy changes in a difference-in-differences framework to analyze retention effects. For example, Ferreira et al. (2018) evaluate a 2010 policy change in Colombia that allowed schools to increase their retention rates above 5%, which increased Spanish test scores but had no effect on math scores. Battistin and Schizzerotto (2019) study an education reform in Italy that changed promotion criteria in upper secondary schools. They exploit geographical variation in the reform's implementation and find heterogeneous

[7] They build on previous work developing such models in Cunha et al. (2006), Cunha et al. (2010), and Agostinelli and Wiswall (2016).

[8] A study by Dong (2010) also uses the ECLS-K to analyze the impact of kindergarten grade retention. She implements a control function estimator that jointly models the choice of enrolling in a school that allows kindergarten retention, the decision for a child to repeat kindergarten, and academic performance in subsequent grades (through grade three). She finds that retention improves academic performance, but the positive effects diminish from 1st to the 3rd grades. Her study compares retained and non-retained students holding grade constant.

impacts, with students in lower educational tracks experiencing negative impacts.

Existing studies typically do not control for dynamic selection bias arising from school dropout, a key concern when evaluating retention effects for older-age youth. Also, most studies focus on test score and education impacts without considering impacts on lifetime earnings or the question of optimal retention policy design. The modeling framework that we develop and implement in this paper is useful for addressing both of these concerns.

3 The Portuguese Education System

3.1 Organization and Governance

The public education system in Portugal is divided into three tiers: preschool education (from ages three to five), basic education (grades 1 through 9), and secondary education (grades 10 through 12). Public education is free and universally available from the age of five. For the cohort we study, school attendance was mandatory for nine years or until age 15. In 2009, attendance through age 18 became mandatory with the passage of Law no. 85/2009, but the first affected cohort entered secondary school in 2012/2013, four years after our cohort was in the 9th grade.

Basic education has a common curriculum, while secondary education provides students with different pathways to match their vocational interests and/or to prepare them for post-secondary studies. Admission to alternative secondary tracks is open to everyone. Students may choose from a general track, oriented towards postsecondary studies, and a vocational track. Access to higher education is through competitive national exams taken in the 12th grade. In principle, all students can take the exams and attend university, but in practice few students outside of the general track take them. Our data cover students pursuing the general track, which includes four subtracks with different course and examination requirements: science and technology, socioeconomic science, languages and humanities, and visual arts.

3.2 Grade Retention in Secondary School

Retention depends on course marks and national exam scores. Course marks range from 0 to 20, and, in grades 10 and 11, a student passes a subject if the mark exceeds 10. A student who passes is promoted, while a student who fails is retained. At the end of the 12th grade, students take the national examinations in subjects that are required by the subtrack in which they are enrolled. In subjects covered by the national exams, the final score is a weighted average of the internal mark (70%) and the exam score (30%), which is rescaled to be between 0 and 20. A student graduates from a 12th grade course if this weighted average rounds up to 10 (≥ 9.5). Students whose score falls below 9.5 are retained except in unusual circumstances. Students may graduate from one course but be retained in another. A student who has not completed one or more courses required by her subtrack does not obtain a degree.

We focus our analysis on learning in math and Portuguese. Portuguese is a required subject for all students, but math is required only for students in the science and technology and socio-economic science subtracks. We pool students in these two subtracks into a single category, which we call STEM. Students in the languages and humanities and visual arts subtracks are grouped together into a non-STEM category. STEM students constitute 76% of our sample.

National examinations at the end of secondary school serve the dual purpose of secondary school exit exams and college entrance exams. The average student fails the math exam and narrowly passes the Portuguese exam (Table 2). Course marks, which are issued before students take the national exams, are more lenient, thereby raising many students' weighted average above the minimum passing threshold. Grading standards can vary across schools, but the national exam scoring occurs outside the students' school region and does not depend on local factors. As described below, our estimation strategy exploits municipality-level variation in course grading criteria as a source of exogenous retention rate variation.

4 Data

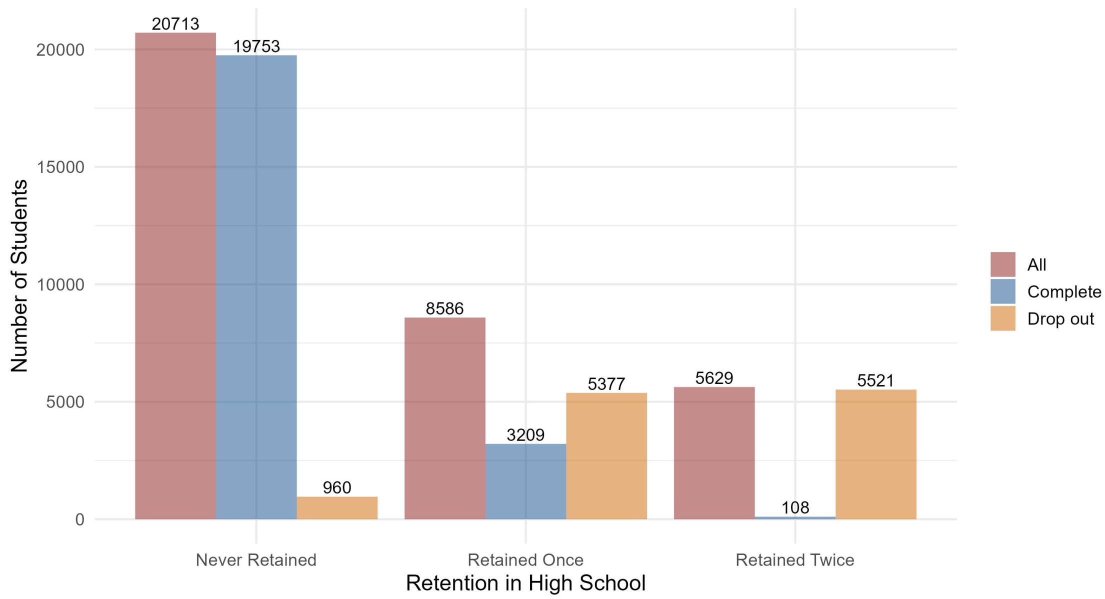
We use an administrative data set obtained from the Portuguese Ministry of Education (MISI) that tracks all public school students during the academic years 2008/09 - 2013/14. It follows 51,119 individuals enrolled in the 9th grade in 2008/09 who advance to the tenth grade in general (not vocational) secondary schools.[9] The data record the schooling trajectory of each student including the sequence of retention, promotion, and dropout decisions, scores on math and Portuguese national exams, and some limited information on course marks. Our structural model will not incorporate course marks, which are not available in the 10th and 11th grades, but we will use marks and exam scores to construct the weighted score that is used, in part, to determine passing in the 12th grade. As previously described, we validate the model by benchmarking model predictions against the RDD estimates. Some observations lack complete records, but the panel is of high quality. After discarding observations with missing data, we end up with a sample of 34,928 individuals, representing 68.3% of the original sample: 28,409 who enter secondary school at the modal age, 4,853 lagging behind one year for their age, and 1,666 lagging behind two or more years.[10]

Figure 1 plots the number of students retained during secondary school, broken down by whether a student graduates or drops out. 59.3% are never retained, 24.6% are retained once, while the remaining 16.1% of students are retained two or more years. A total of 11,858 students drop out, of whom 45.3% were retained one year and 46.6% were retained two or more years during secondary school. Retention rates vary considerably by grade level and course. 20.6% of students who enroll in the tenth grade are retained in that grade, whereas only 9.3% of students who enroll in the eleventh grade are retained. Retention in the 12th grade is subject-specific: 25.5% (8.5%) of students enrolled in math (Portuguese)

[9]The initial data set comprises all individuals, regardless of age, enrolled in a public school in the 9th grade in the academic year 2008/09 representing a total of 82,412 individuals. In the following year, students may be repeating 9th grade, be in the 10th grade general track or 10th grade vocational track, or have dropped out. We analyze the 51,119 students who transition to the 10th grade general track.

[10]Table B-1 in Appendix B compares the full sample and our estimation sample using means and standard deviations of variables that are rarely missing. The estimation sample is slightly positively selected on ninth grade test scores and family income.

Figure 1: Retention and Secondary School Completion



are retained.

The data provide information on individual and family characteristics for each student, including their grade, the subtrack in which they are enrolled, their gender, age, and nationality. We also have information on family background, including parental education, whether the family's income is low enough to qualify for a public subsidy, as well as information on the student's school, classes, and teachers.[11] Tables 1 and 2 present descriptive statistics from the estimation sample for all variables used in our analysis, differentiated by whether students are retained (in any year) or promoted (in all years). Girls are slightly overrepresented (55%) among students but are underrepresented among retained students. One-third of students in the sample are low-income, and low-income students are more likely to be retained. Among retained students, 53% have mothers with less than a basic education (less than nine years) and only 10% have mothers with more than a secondary school education. These figures are 40% and 24% for students who were never retained.

[11]Students attend all classes with the same set of peers, so class size is the same in both Portuguese and math.

Table 1: Descriptive Statistics - Demographics

Variable	Category	All	Never Retained	Retained
Gender	Female	0.55	0.60	0.48
Age at Secondary	16 or Under	0.81	0.89	0.70
School Entry	17	0.14	0.09	0.22
	18 or Older	0.05	0.02	0.09
Family	Low	0.33	0.28	0.38
Income	High	0.67	0.72	0.62
Mother's Ed	Less than Basic/Unknown	0.45	0.40	0.53
	Basic	0.18	0.17	0.19
	Sec. School	0.19	0.20	0.18
	More than Sec. School	0.18	0.24	0.10
Geography	Rural	0.21	0.21	0.21
	Semiurban	0.33	0.35	0.30
	Urban	0.46	0.45	0.49

We classify family income as low if the family qualifies for a public subsidy. Basic education corresponds to nine years of schooling.

Table 2 shows that students who experience retention have below-average test scores in both the 9th and the 12th grades, and the gap between retained and nonretained students widens between these grades. There is little difference in teacher characteristics and class size between retained and promoted students. The third panel of Table 2 shows the means of two variables that we will use as exclusion restrictions in our model's estimation. Dropouts earn 134 fewer euros per month on average than secondary school graduates. The historical retention rate in the concelho (municipality) averages 21.8% across all grades and concelhos.

Table 2: Descriptive Statistics - Test Scores, Teacher Char., and Exclusion Restrictions

Variable	Category	All	Never Retained	Retained
Test Scores				
Math	9th	62.40	69.17	52.53
		(20.38)	(19.45)	(17.45)
	12th	48.12	57.25	26.17
		(23.11)	(19.1)	(16.08)
Portuguese	9th	59.90	64.98	52.49
		(14.76)	(13.76)	(12.92)
	12th	50.75	55.41	40.44
		(16.81)	(15.58)	(14.7)
Teacher Characteristics				
Female Teacher	Math	0.75	0.74	0.75
		(0.43)	(0.44)	(0.43)
	Portuguese	0.86	0.86	0.86
		(0.35)	(0.35)	(0.35)
Teacher Age	Math	46.59	47.08	46.01
		(7.75)	(7.55)	(7.94)
	Portuguese	48.71	48.91	48.42
		(6.16)	(6.09)	(6.25)
Exclusion Restrictions				
Local Wage Difference	Dropout - Grad	-133.98	-128.19	-142.42
(Euros/mo)		(63.54)	(60.54)	(66.78)
Concelho Retention Rate ($t-3$)		21.77	21.21	22.58
		(6.06)	(5.96)	(6.12)

The first observed 12th grade test score is used to compute the mean for retained students.

5 Model

We develop a model of knowledge accumulation in math and Portuguese, grade retention, and secondary school dropout. In each year, students realize one of two potential outcomes for knowledge accumulation that depend on whether the student was retained or not in the prior year (denoted by $R_{i,t-1} \in \{1, 0\}$). For a given subject ($S \in \{M, P\}$), the two potential outcome equations are:

$$K^S_{i,t}(1) = \gamma^S_1 K^S_{i,t-1} + I^{S\,\prime}_{i,t} \beta^S_1 + \varepsilon^S_{i,t}(1) \;, \tag{1}$$

$$K^S_{i,t}(0) = \gamma^S_0 K^S_{i,t-1} + I^{S\,\prime}_{i,t} \beta^S_0 + \varepsilon^S_{i,t}(0) \;, \tag{2}$$

where $I_{i,t}$ denotes school characteristics and family background variables. The data do not include detailed information on family input choices (e.g. homework help), so including family background covariates can help to control for these inputs. Our value-added formulation makes the standard assumption that lagged knowledge $K^S_{i,t-1}$ is a sufficient statistic for prior inputs. The model allows the coefficients of the knowledge production function for retained students to differ from those of promoted students, because retained students attend the same grade for a second time while promoted students are exposed to new curricula.

5.1 Dropout and Grade Retention

We model retention and dropout using the following threshold-crossing equations:

$$R_{i,t} = \mathbb{1}(\lambda_0 + \lambda_1 K^M_{i,t} + \lambda_2 K^P_{i,t} + \lambda_3 Z^R_{i,t} + \lambda_4 I_{i,t} + \nu_{i,t} > 0) \;, \tag{3}$$

$$D_{i,t} = \mathbb{1}(\delta_0 + \delta_1 K^M_{i,t-1} + \delta_2 K^P_{i,t-1} + \delta_3 Z^D_{i,t} + \delta_4 \underbrace{\sum_{k>0} R_{i,t-k}}_{\text{Prior Retentions}} + \delta_5 I_{i,t-1} + \eta_{i,t} > 0) \;. \tag{4}$$

Retention in year t depends on knowledge levels in math ($K^M_{i,t}$) and Portuguese ($K^P_{i,t}$), on school and family characteristics ($I_{i,t}$), and on an exclusion restriction, $Z^R_{i,t}$. Dropout in year t depends on knowledge and investment in the prior year, an exclusion restriction representing contemporaneous labor market opportunities, $Z^D_{i,t}$, and the sum of prior retentions in secondary school, $\sum_{k>0} R_{i,t-k}$. The direct effect of retention on the decision to drop out is captured by δ_4. If $\delta_4 > 0$, then students are discouraged by past retentions. Retention could also indirectly influence dropout through its causal effects on knowledge in the next period, governed by equations (1) and (2). If greater knowledge reduces the probability of dropping out ($\delta_1, \delta_2 < 0$) and retention raises knowledge for some students, then this may offset some of the direct effects of retention on dropout.[12]

In estimation, we incorporate a few modifications to equations (1) - (4) so that the estimated model better fits the data. First, we allow test scores to also depend on the teacher characteristics shown in Table 2. Second, dropout and retention rates vary by grade and subject, so we allow the intercepts in equations (3) and (4) vary by grade and subject. Third, retention in Portuguese in grade 12 has both a separate intercept and slope and does not depend on math knowledge, and retention in math in grade 12 is treated analogously. (Recall that retention in 12th grade is grade-specific). Finally, we allow for differences between STEM and non-STEM students by having separate non-STEM track intercepts in equations (3) and (4).

5.2 Distribution of Unobservables

The model allows for correlation between the unobserved shocks in the grade retention, dropout, and test score equations. We assume that the vector of model shocks is jointly normally distributed as follows:

$$\begin{pmatrix} \varepsilon^M_{i,t}(0) \\ \varepsilon^M_{i,t}(1) \\ \varepsilon^P_{i,t}(0) \\ \varepsilon^P_{i,t}(1) \\ \eta_{i,t+1} \\ \nu_{i,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma^2_M(0) & - & \sigma_{MP}(0) & - & \sigma_{M\eta}(0) & \sigma_{M\nu}(0) \\ & \sigma^2_M(1) & - & \sigma_{MP}(1) & \sigma_{M\eta}(1) & \sigma_{M\nu}(1) \\ & & \sigma^2_P(0) & - & \sigma_{P\eta}(0) & \sigma_{P\nu}(0) \\ & & & \sigma^2_P(1) & \sigma_{P\eta}(1) & \sigma_{P\nu}(1) \\ & & & & 1 & \sigma_{\eta\nu} \\ & & & & & 1 \end{pmatrix} \right) \tag{5}$$

[12] Appendix D shows how the threshold-crossing equation for dropout in (4) can be derived from the optimization problem of a student who compares expected earnings streams conditional on dropping out or staying in school.

We further assume that shocks that are not contemporaneous are independent:

$$
\begin{aligned}
\varepsilon^M_{i,t}(0) &\perp\!\!\!\perp \varepsilon^M_{i,t+j}(1), \varepsilon^P_{i,t+j}(0), \varepsilon^P_{i,t+j}(1), \eta_{i,t+1+j}, \nu_{i,t+j} \ , \\
\varepsilon^M_{i,t}(1) &\perp\!\!\!\perp \varepsilon^P_{i,t+j}(0), \varepsilon^P_{i,t+j}(1), \eta_{i,t+1+j}, \nu_{i,t+j} \ , \\
\varepsilon^P_{i,t}(0) &\perp\!\!\!\perp \varepsilon^P_{i,t+j}(1), \eta_{i,t+1+j}, \nu_{i,t+j} \ , \\
\varepsilon^P_{i,t}(1) &\perp\!\!\!\perp \eta_{i,t+1+j}, \nu_{i,t+j} \ , \\
\eta_{i,t+1} &\perp\!\!\!\perp \nu_{i,t+j} \ ,
\end{aligned}
$$

for $j \neq 0$.

As in standard versions of the generalized Roy Model, certain covariances are not identified. An individual is never retained and promoted in the same year, so we cannot identify the covariance between the shocks in the knowledge accumulation equations across these two states. But, these covariances are not necessary for our counterfactual simulations. Specifically, we use the estimated model to analyze policies that alter retention rates to understand the effects on dropout and test score distributions. These counterfactual distributions depend in part on the covariances in the final column of the covariance matrix in (5), which are identified.

5.3 Model Implications

In equations (1) and (2), there are multiple ways in which retention may influence test scores. To see this, consider two grade progression trajectories: $10-11-12$ and $10-10-11-12$. If student i follows the trajectory $10-11-12$, we can use back-substitution to write the observed knowledge measure in grade 12 as a function of observed test scores and investment and unobserved shocks as follows:

$$K_{i,t+3} = {\gamma_0}^3 K_{i,t} + ({\gamma_0}^2 I_{i,t+1} + \gamma_0 I_{i,t+2} + I_{i,t+3})'\beta_0 + \varepsilon_{i,t+3}(0) + \gamma_0 \varepsilon_{i,t+2}(0) + {\gamma_0}^2 \varepsilon_{i,t+1}(0) \quad (6)$$

If the student had instead been retained in the 10th grade and followed the trajectory $10-10-11-12$, her grade 12 knowledge would be as follows:

$$\begin{aligned}K_{i,t+4} = {\gamma_0}^3\gamma_1 K_{i,t} + ({\gamma_0}^2\gamma_1 I_{i,t+1} + \gamma_0 I_{i,t+3} + I_{i,t+4})'\beta_0 + {\gamma_0}^2 I_{i,t+2}\beta_1 + \varepsilon_{i,t+4}(0) + \\ \gamma_0\varepsilon_{i,t+3}(0) + {\gamma_0}^2\varepsilon_{i,t+2}(1) + {\gamma_0}^2\gamma_1\varepsilon_{i,t+1}(0)\end{aligned} \tag{7}$$

A comparison of (6) and (7) shows that retention can lead to different 12th grade test scores, because of the additional year of investment, $I_{i,t+4}$, differences in the productivity of investment in the retained year $I'_{i,t+2}(\beta_1 - \beta_0)$, an additional year of skill depreciation, potentially different rates of depreciation following the retained year, $\gamma_1 \neq \gamma_0$, and different shocks. Our model is flexible enough to allow for all of these mechanisms and differentiate between them.

6 Estimation and Identification

6.1 Estimation

We estimate the model using simulated method of moments (SMM) and an unconditional simulation approach (Gourieroux et al., 1996), which means that we simulate each student's path of educational outcomes throughout secondary school starting from their initial conditions at the end of the ninth grade. Model parameters are chosen to minimize the weighted distance between the model-simulated outcomes and the corresponding data moments.

Table 3 lists common trajectories for Portuguese students who are enrolled in the STEM and non-STEM tracks. As previously noted, students are sometimes retained in one subject but not the other. We define a variable, h_i, that represents an individual's history, which is the Cartesian product of an individual's math path and Portuguese path, $h_i \equiv PathMath_i \times PathPT_i$. We denote the set of all possible histories by $\mathcal{H}$. Table 3 lists the most common histories in the data and the corresponding observed outcomes

To simplify notation, let θ denote the vector of model parameters, $y_{i,t} = (K^M_{i,t}, K^P_{i,t}, h_i)$ the vector of endogenous variables, $z_i = \{K^{9,M}_i, K^{9,P}_i, I_{i,t}, Z^R_{i,t}, Z^D_{i,t}, \}^T_{t=1}$ the vector of exogenous variables, and $\epsilon_{\mathbf{i}} := \{\{\epsilon^s_{i,t}\}^T_{t=1}\}^S_{s=1} = \{\{\varepsilon^{M,s}_{i,t}(0), \varepsilon^{M,s}_{i,t}(1), \varepsilon^{P,s}_{i,t}(0), \varepsilon^{P,s}_{i,t}(1), \eta^s_{i,t}, \nu^s_{i,t}\}^T_{t=1}\}^S_{s=1}$ the set

Table 3: Common Secondary School Histories and Test Score Outcomes

STEM		Non-STEM	
History	Observed Scores	History	Observed Scores
10-11-12X10-11-12	$K^M_{i,t+3}, K^P_{i,t+3}$	10-11-12	$K^P_{i,t+3}$
10-10dX10-10d		10-10-11-12	$K^P_{i,t+4}$
10-11-12-12-12dX10-11-12	$K^M_{i,t+3}, K^P_{i,t+3}, K^M_{i,t+4}$	10-11-12-12d	$K^P_{i,t+3}$
10-11-12-12X10-11-12	$K^M_{i,t+3}, K^P_{i,t+3}, K^M_{i,t+4}$	10-11-12-12-12d	$K^P_{i,t+3}, K^P_{i,t+4}$
10-11-12dX10-11-12d		10-11-12-12	$K^P_{i,t+3}, K^P_{i,t+4}$
10-10-10dX10-10-10d		10-11-11-12	$K^P_{i,t+4}$
10-11-12-12dX10-11-12	$K^M_{i,t+3}, K^P_{i,t+3}$	10-10-11-12-12d	$K^P_{i,t+4}$
10-11-11dX10-11-11d		10-10-11-11-12	
10-10-11-12X10-10-11-12	$K^M_{i,t+4}, K^P_{i,t+4}$	10-11-11-12-12d	$K^P_{i,t+4}$
10-10-11-11dX10-10-11-11d			
10-11-11-11dX10-11-11-11d			
10-11dX10-11d			
10-11-11-12X10-11-11-12	$K^M_{i,t+4}, K^P_{i,t+4}$		
10-11-12-12X10-11-12-12	$K^M_{i,t+3}, K^P_{i,t+3}, K^M_{i,t+4}, K^P_{i,t+4}$		
10-11-12-12-12dX10-11-12-12-12d	$K^M_{i,t+3}, K^P_{i,t+3}, K^M_{i,t+4}, K^P_{i,t+4}$		

The table lists common histories and associated test scores in the twelfth grade that can be observed in the data.

of errors for all S simulations and T time periods. A history h_i encodes all dropout and retention decisions, so it is not necessary to include $R_{i,t}$ and $D_{i,t}$ in $y_{i,t}$.

We simulate the full path of endogenous variables as a function of exogenous variables and shocks. A given simulation, s, can be written as $y^s_{i,t} = r(y^s_{i,t-1}(\theta), z_i, \epsilon^s_i)$ where $y^s_{i,t-1}(\theta)$ is a simulated value that depends on the parameter vector θ as well as prior realizations of z_i and ϵ^s_i. A path simulation allows us to write $y^s_{i,t}$ as a function of only the exogenous variables, z_i, the shocks, $\epsilon^{\mathbf{s}}_{\mathbf{i}}$, and the initial value of the process, $y_{i,0}$: $y^s_{i,t} = r(y_{i,0}, z_{i,1}, \ldots, z_{i,t}, \epsilon^s_{i,1}, \ldots, \epsilon^s_{i,t})$.

The SMM estimator minimizes the following objective function:

$$J(\theta) = \left\{ \frac{1}{N} \sum_{i=1}^{N} z_i [F(y_i) - f(z_i, \epsilon_{\mathbf{i}}, y_{i,0}; \theta)] \right\}' \Omega^*(\theta) \left\{ \frac{1}{N} \sum_{i=1}^{N} z_i [F(y_i) - f(z_i, \epsilon_{\mathbf{i}}, y_{i,0}; \theta)] \right\}, \tag{8}$$

where $F(y_i)$ is a function only of the data and $f(z_i, \epsilon_{\mathbf{i}}, y_{i,0}; \theta)$ is its corresponding simulated value. $f(z_i, \epsilon_{\mathbf{i}}, y_{i,0}; \theta)$ is an unbiased simulator of $F(y_i)$, meaning that $\mathbb{E}[f(z_i, \epsilon_{\mathbf{i}}, y_{i,0}; \theta) \mid z_i] = \mathbb{E}[F(y_i) \mid z_i]$.

We target three types of moments corresponding to three broad groups of dependent variables. These dependent variables are an indicator for whether individual i has a particular history, $\mathbb{1}_{(h_i = h)}$, the 12th grade test score for histories in which the student reaches the end of the 12th grade, $\mathbb{1}_{(h_i=h)}K_{i,t}$, and squared test scores for the same histories, $\mathbb{1}_{(h_i=h)}(K_{i,t})^2$. The moments that we target, denoted $m_i = z_i \cdot F(y_i)$, therefore represent covariances between exogenous variables and a student's history, covariances between exogenous variables and observed test scores, and test score second moments. Appendix A shows the full list of moments used in estimation.

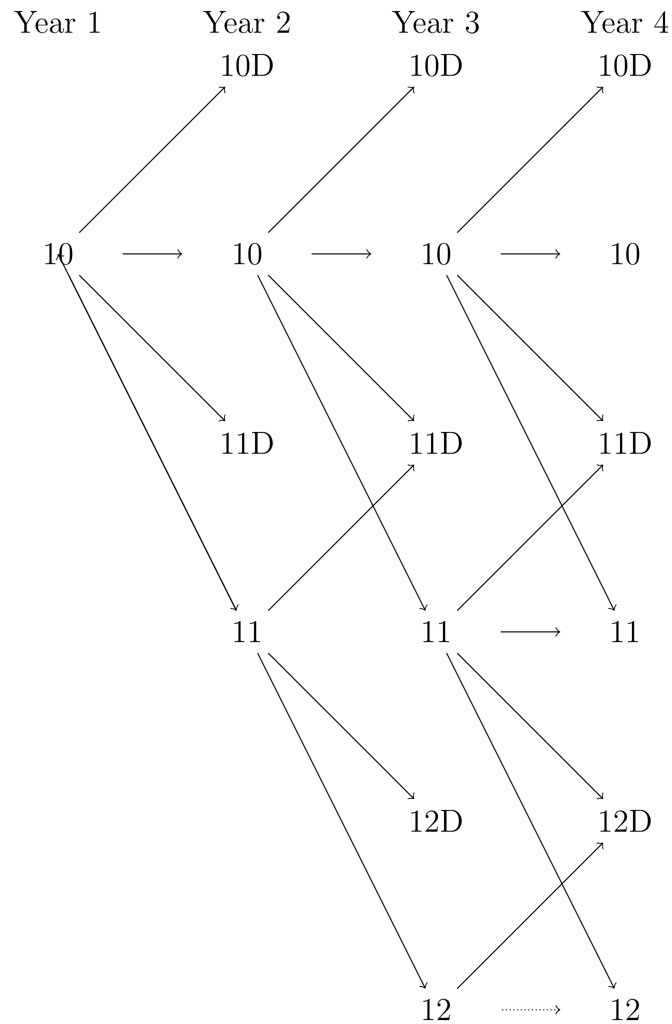
We use the optimal weight matrix, $\Omega^*(\theta)$, when minimizing $J(\theta)$ (Gourieroux et al., 1996). In practice, we first optimize $J(\theta)$ with the weight matrix given by the inverse of variances of the sample moments to obtain an estimate of θ and then calculate an estimate of the optimal weight matrix at the initial solution and re-optimize with the new weight matrix. Because $J(\theta)$ is discontinuous in θ, we use a logit smoother to simulate the binary dependent variables. This replaces the dependent variables $\mathbb{1}_{(h_i=h)}$ with a smooth probability. The smoothing parameter is set to $\lambda = 0.10$. Tables B-2 and B-3 in Appendix B compare the fit of the model using the smoothed and crude simulators and show that both simulators predict nearly identical proportions of individuals with each history.

6.2 Identification

Figure 2 provides a visual depiction of the possible paths students can take through secondary school. Numbers represent grade levels, and those followed by a "D" indicate a path where a student dropped out after enrolling in that grade. The retention and dropout equations govern which edges students travel along as they move rightward from each node.

We now describe how we identify the unobservable shock distribution parameters shown in (5). As previously noted, covariances between the test score shocks in the retained and promoted states are not identified as students are never observed in both states. The remaining parameters of (5) can be identified from our assumption that the error distribution is jointly normal, but exclusion restrictions provide additional sources of identification that do not rely on functional form. Also, in theory, the error distribution could vary at each node in Figure 2, but because test scores are not observed in every grade, we constrain the covariance matrix for unobserved shocks to be equal across grades. Appendix E shows how

Figure 2: Possible Secondary School Paths Over Four Years



The figure depicts the set of possible paths for a particular subject through four years of secondary school. In year 10, the individual is in 10th grade. Moves rightward or up and to the right in subsequent years indicate retentions. Moves downward and to the right indicate promotions. Any node in which a number is followed by a "D" indicates dropout. The line between 12 in year 3 and year 4 is dotted, because only some students will be retained.

these covariances can be identified from regression models that include control functions, estimated on different subsets of students with the same history.

The average retention rate in the concelho three years prior serves as a time-varying exclusion restriction for the retention equation. Our theory for why this variable affects retention without directly affecting dropout or test scores is as follows. Retention in secondary school is based, in part, on course marks and a high retention rate three years prior may, to some extent, reflect that teachers applied harsher grading standards.[13] However, because marks are a relative measure of skill, they do not directly affect knowledge production or dropout. We test these suppositions by estimating regressions using data from the 12th grade, the only grade in which we observe course marks. Table B-4 in Appendix B provides evidence for the first supposition by showing that course marks in the 12th grade are lower in districts with high values of the exclusion restriction even conditional on test scores. Table B-5 then provides evidence for the second supposition by showing that marks do not predict dropout conditional on test scores. Although course marks are predictive of the decision to drop out in the 12th grade, their partial correlation with dropout vanishes once scores on the national exam are included.[14] This evidence suggests that the three-year lagged retention rate may be plausibly excluded from the other model equations, but in Appendix B we examine the robustness of our estimation results to an alternative exclusion restriction.

We use variation in concelho-level differences in wages between secondary school dropouts and graduates as an exclusion restriction that affects the dropout decision. Local labor market conditions, which have been used in the context of employment and child care decisions, would be expected to influence dropout behavior by altering the opportunity cost of remaining in school (Åslund and Rooth, 2007; Bernal, 2008).

7 Empirical results

7.1 Estimated Parameters

Table 4 presents the parameter estimates for the knowledge accumulation equations for math and Portuguese (equations (1) and (2)). The estimated parameters show that, for promoted

[13] Teachers in Portugal remain with a student cohort as they advance through secondary school and, for this reason, the same teachers will teach 10th grade every three years. Therefore, students in our cohort will primarily have the same set of instructors who were responsible for producing the retention rate in the concelho three years prior. To the extent that teachers have heterogeneous grading standards, students will be more or less likely to experience retention depending on the teacher to whom they are assigned. Only in exceptional circumstances do students have a choice over concelhos in which to attend school.

[14] We estimate these probit models on the subsample of students for whom we observe test scores, course marks, and a dropout decision, namely those who are retained in the 12th grade and then have a decision regarding whether to repeat the 12th grade or drop out. Since dropout causes a sample selection problem, it is not straightforward to test whether the prior year's grades affect test scores in the same subsample.

students, being from a lower income family reduces learning in math and Portuguese, while having a more educated mother increases learning. Teacher experience increases test scores. For promoted students, having a female teacher reduces performance in math but raises it in Portuguese. The effect of retention on knowledge in subject S in the year that the retention occurs for the subgroup of students who are retained and who decide to stay in school is given by:

$$\Delta_{TT} = (\gamma_1 - \gamma_0)K^S_{i,t-1} + (\beta_1 - \beta_0)I_{i,t} + E(\varepsilon^S_{i,t}(1) - \varepsilon^S_{i,t}(0)|K^S_{i,t-1}, I_{i,t}, R_{it} = 1, D_{it} = 0)\ .$$

The expression shows that heterogeneity in the effects of retention by initial skill, $K^S_{i,t-1}$, is governed by $\gamma_1 - \gamma_0$. We find that $\gamma_1 < \gamma_0$ for both math and Portuguese, which means that retention has more positive effects for less skilled students. We also find that the intercept in β_1 is greater than the the intercept in β_0 for math, but that the opposite is true for Portuguese. For math, most of retention's positive effects on test scores will be driven by the difference in these constants, reflecting the differential productivity of learning in the year the student is retained.

Table 5 shows the dropout parameter estimates. The dropout model allows for grade-specific intercepts but, for reasons of parsimony, assumes that other coefficients do not differ by grade. Students with higher test scores are significantly less likely to drop out as are students from higher income families and those with more educated mothers. Also, girls are less likely to drop out. Table 6 shows the estimated retention model coefficients. Again, the intercepts are allowed to vary across grades; other coefficients are equal across grades and subjects, except for the 12th grade slope coefficients (because 12th grade retention in math does not depend on Portuguese skills and vice versa for 12th grade retention in Portuguese). Higher test scores reduce the retention probability in all grades and subjects. Students in larger classes and males are more likely to be retained. The exclusion restrictions positively affect retention and dropout.

Table 7 shows the estimated covariance parameters for the error terms in the test score, dropout, and retention equations. The estimated covariance between the shocks in the

Table 4: Estimated Value-Added Equation Parameters

	Math				*Portuguese*			
	First time		Repeating		First time		Repeating	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Lagged Score, $K_{i,t-1}$	0.973	0.004	0.680	0.019	0.837	0.003	0.514	0.029
Constant	-8.528	0.592	14.526	2.381	3.203	0.503	-1.962	6.225
Relative Age at Gr 10	-3.744	0.196	-0.489	0.547	-1.855	0.081	1.693	0.614
Female Teacher	-0.799	0.159	1.106	0.729	0.385	0.151	-0.367	1.894
Teacher Age	0.366	0.089	-0.565	0.423	0.454	0.087	2.110	1.194
Class size	-0.142	0.153	0.099	0.528	-0.078	0.089	1.873	0.978
Low SES	-1.038	0.135	0.194	0.549	-0.167	0.085	3.588	1.089
Mother Basic Educ	0.166	0.147	0.876	0.765	-0.080	0.091	-1.372	1.534
Mother Sec. Educ	0.567	0.141	0.697	0.735	0.186	0.09	-2.922	1.654
Mother > Sec. Educ	2.344	0.140	2.104	0.941	1.411	0.095	-1.936	2.227
Semiurban	0.200	0.140	-1.261	0.774	0.178	0.092	1.014	1.678
Urban	-0.114	0.142	0.402	0.761	-0.079	0.090	-0.001	1.511
Female	1.029	0.099	2.186	0.560	0.779	0.066	0.623	1.154

The table presents parameter estimates for the four value-added equations. The omitted education category is less than basic. Class size is measured in tens of students, and teacher age is measured in tens of years.

retention and dropout equations is fairly high (0.635). There is positive sorting on math test score gains into retention, $\sigma_{M\nu}(1) - \sigma_{M\nu}(0) = 4.789$, but negative sorting on Portuguese test score gains, $\sigma_{P\nu}(1) - \sigma_{P\nu}(0) = -8.221$.[15]

7.2 Model Fit

We target moments for a large number of histories in estimation (Table 3), and the model is over-identified. Tables B-2 and B-3 in Appendix B show the model fit for the proportion of students with each history. The tables show that the model does a good job of matching the most common histories for STEM students, although it slightly underpredicts the fraction of individuals with less common histories. For example, 37.9% of students follow the STEM history 10-11-12X10-11-12 in the data and 41.0% in model simulations, while 6.6% of students follow history 10-10dX10-10d in the data and 7.4% in model simulations. The third column

[15]Appendix Tables B-6 through B-9 we explore the robustness of our results to an alternative exclusion restriction, retention rates at other schools in the same concelho. This approach produces similar parameter estimates as our main specification.

Table 5: Dropout Equation Parameters

	Estimate	SE
Intercept - 10th grade	1.294	0.106
Intercept - 11th grade	1.338	0.106
Intercept - 12th grade	1.013	0.096
Math	-3.212	0.083
Portuguese	-2.333	0.118
Local Income: Dropout - Graduate	0.037	0.015
Relative Age at Gr 10	0.369	0.018
Yrs Retained in Sec. School	-0.116	0.038
Class size	0.017	0.023
Low SES	0.053	0.021
Mother Basic Educ 1	-0.074	0.026
Mother Sec. Educ 2	-0.138	0.027
Mother > Sec. Educ 3	-0.108	0.035
Semiurban	0.052	0.028
Urban	0.113	0.028
Female	-0.048	0.021
Non-STEM (12th grade)	-1.027	0.064

The omitted education category is less than basic. The coefficients and standard errors on math and Portuguese knowledge have been scaled so that they represent the effects of a 100-point increase in these scores. Class size is measured in tens of students. Income is measured in 100's of Euros per month.

shows that the smooth and crude simulators produce very similar moments. Table B-3 shows that the model is somewhat less able to match infrequent non-STEM histories: 2.6% (0.9%) of students in the data (simulation) follow the 10-10-11-12 path, and the simulation underpredicts other infrequent histories.

7.3 Retention Impacts

We next analyze how grade retention influences students by simulating four separate policies. First, we evaluate the effects of the status quo retention policy on the subgroup of students who are retained. To do so, we use the estimated model to simulate outcomes and then compare them to those obtained from a simulation in which all students are promoted at the end of each grade. This yields an estimate of the average effect of treatment on the treated

Table 6: Retention Equation Parameters

	Estimate	SE
Intercept - 10th grade	3.455	0.126
Intercept - 11th grade	2.216	0.029
Intercept - 12th grade, Math	4.198	0.170
Intercept - 12th grade, Portuguese	1.380	0.112
Slope - Math, 10/11th grades	-4.487	0.144
Slope - Math, 12th grade	-16.47	0.531
Slope - Portuguese, 10/11th grades	-4.523	0.153
Slope - Portuguese, 12th grade	-8.460	0.199
Historical Retention Rate	1.184	0.149
Relative Age at Gr 10	0.186	0.018
Class size	0.107	0.022
Low SES	-0.080	0.020
Mother Basic Educ	-0.032	0.025
Mother Sec. Educ	-0.069	0.025
Mother > Sec. Educ	0.004	0.031
Semiurban	-0.045	0.026
Urban	0.093	0.025
Female	-0.046	0.020
Non-STEM	-3.168	0.077

Retention is grade-specific in grades 10 and 11 but grade-subject-specific in grade 12. Retention in grade 12 in Math depends on math scores but not Portuguese scores and vice versa for retention in grade 12 in Portuguese. The omitted education category is less than basic. The coefficients and standard errors on math and Portuguese knowledge have been scaled so that they represent the effects of a 100-point increase in these scores. The historical retention rate is measured on a 0-1 scale.

(TT). The remaining three policies are used to obtain estimates of average treatment effects (ATEs) that correspond to retaining all students in a particular grade. We estimate the ATE of retention in grade g by retaining all students once in grade g and promoting them in other grades and comparing this outcome to a simulation in which all students are promoted in all years.

Table 8 shows the estimated test score impacts for the first policy simulation for the subgroup of students who experience one retention and graduate in four years (instead of three). The status quo policy causes, on average, a 0.46 sd increase in math scores and a 0.17 decrease in Portuguese scores. The rightmost table column indicates that 7% of secondary

Table 7: Estimated Covariance Matrix

$\varepsilon^M(0)$	$\varepsilon^M(1)$	$\varepsilon^P(0)$	$\varepsilon^P(1)$	Dropout Shock	Retention Shock
115.839	0	54.047	0	1.975	1.542
(1.608)	–	(1.448)	–	(0.359)	(0.318)
...	123.447	0	-78.868	6.628	6.331
...	(5.046)	–	(4.515)	(1.041)	(0.425)
...	...	72.365	0	1.043	4.176
...	...	(0.766)	–	(0.305)	(0.271)
...	...	...	50.387	-4.235	-4.045
...	...	...	(4)	(1.821)	(0.409)
...	...	...	...	1	0.635
...	...	...	...	...	(0.021)
...	...	...	...	...	1

school students are retained and graduate in exactly four years. The ATEs in which all students are retained in grades 10, 11, and 12 are negative for both math and Portuguese. The fact that TT > ATE indicates that there is positive sorting into retention on the basis of test score gains for students who graduate in four years.

We also consider the effect of eliminating grade retention on the probability of dropping out. In the data, 34% of Portuguese secondary school students and fully 77% of retained students drop out. Set against these numbers, the 7 pp point increase in dropout caused by the status quo retention policy in Table 9 seems small. Boys are more likely to drop out than girls, but the effect of the status quo policy on dropout does not differ by gender. The effects on dropout are smaller than the overall frequency of dropout, because most retained students would drop out even in the absence of being retained. Appendix Table 7 shows that the correlation between shocks to retention and shocks to dropout is 0.635. Retained students also have observable factors, such as low test scores, that make them more likely to drop out.

Table 8: The Effect of Retention on 12th Grade Test Scores

	Math			*Portuguese*			
Policy	Raw	S.D	$TE > 0$	Raw	S.D.	$TE > 0$	Grad in 4 Years
Status Quo (TT)	10.88	0.46	0.82	-2.91	-0.17	0.42	0.07
	(0.70)	(0.03)	(0.01)	(1.10)	(0.07)	(0.04)	
Retain 10	-6.16	-0.26	0.37	-10.8	-0.64	0.21	0.80
	(0.99)	(0.04)	(0.02)	(0.73)	(0.04)	(0.02)	
Retain 11	-3.69	-0.16	0.43	-11.35	-0.68	0.22	0.79
	(0.74)	(0.03)	(0.02)	(0.81)	(0.05)	(0.02)	
Retain 12	-1.14	-0.05	0.47	-12.99	-0.77	0.11	0.75
	(0.63)	(0.03)	(0.02)	(0.84)	(0.05)	(0.02)	

The table shows estimates of test score treatment effects corresponding to four separate retention policies: the status quo retention policy, and policies that retain all students in grades 10, 11, and 12 respectively. The standard deviations on the 12th grade math and Portuguese exams are 23.5 and 16.8 points. The column labeled $TE > 0$ indicates the fraction with a positive treatment effect, while the column labeled Grad in 4 Years indicates the proportion in the simulation who finish secondary school in 4 years. Standard errors, obtained from 200 parametric bootstrap replications, are shown in parentheses.

7.4 Model Validation

As previously described, for 12th grade students who are near the margin of passing or being retained, we can apply a regression discontinuity (RDD) estimator to estimate retention impacts. These RDD estimates are of interest in their own right and serve as a benchmark for evaluating the reliability of our model in simulating retention impacts.

In 12th grade, retention decisions are based on a weighted average of course marks and test scores that we have in our data. The threshold for passing math/Portuguese is 9.5. A student passes if the following weighted average of her course mark and score on the national exam rounds up to 10:

$$Score_i = 0.3 \times Exam_i/5 + 0.7 \times Mark_i \ ,$$

where the rescaled exam score and the course mark both range from 0 to 20. The retention rules are strictly applied, but students have the option of retaking the exam in the summer if they fail the first exam in the spring. Selective exam retaking may cause some manipulation around the threshold, so we estimate a fuzzy regression discontinuity design where the running variable, $Score_i$, is calculated using the first exam score. We subtract 9.5 from $Score_i$

and plot the distribution of scores relative to the threshold for both math and Portuguese in Figure 3. There does not appear to be any evidence of manipulation in the score distribution around the threshold.[16]

Figure 4 plots the probability of dropping out as a function of $Score_i$. Dropout becomes less common in both math and Portuguese as scores approach the passing threshold. There is an apparent discontinuity in dropout probabilities at the passing threshold in Portuguese but not in math.[17] The fuzzy RDD estimates are presented in the third panel of Table 9 together with 95% asymptotic confidence intervals. For students at the margin of retention, retention in math leads to a 12.1 percentage point (pp) increase in the probability of dropping out, while retention in Portuguese leads to 52.5 pp increase in the probability of dropping out. The confidence intervals are fairly wide, [-6.4, 30.6] for math and [31.6 , 73.5] for Portuguese, and only the retention effect in Portuguese is statistically significantly different from zero.[18]

Table 9 also shows the point estimates of the 12th grade retention effects obtained from our model in each subject on the probability of dropping out. In math, the model predictions closely replicate the RDD estimates. The RDD estimate is 0.121 and the model predicts 0.116. For Portuguese, the RDD estimate is 0.525 compared to 0.427 predicted by the model. The confidence intervals for the RD estimates contain our model-simulated estimates for both math and Portuguese.

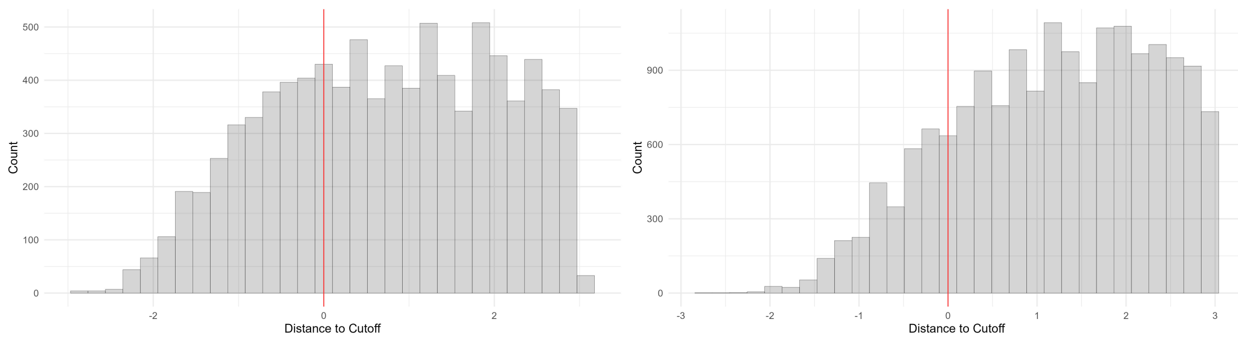
To summarize, model-based simulations reproduce the pattern seen in the RDD estimates that retention in Portuguese has significantly greater effects on dropping out than retention

[16] Figure C-1 in Appendix C examines covariate balance around the threshold. It shows that prior test scores also vary smoothly around the cutoffs for both math and Portuguese.

[17] Appendix Figure C-2 plots the first stage effect of passing the threshold on retention. No students are retained to the right of the thresholds, while 19.1% are retained in math and 13.5% percent are retained in Portuguese directly to the left of the thresholds. All regressions are estimated using local linear regression with an Epanechnikov kernel. The bandwidths, of 0.528 for math and 0.492 for Portuguese, are selected to minimize mean squared error using the direct plug-in approach described in Calonico et al. (2017).

[18] We explore how the RDD results vary with the choice of bandwidth in Appendix C. Because retention and dropout are both deterministic to the right of the passing threshold, the bandwidths selected by the approach described in Calonico et al. (2017), which are used to produce the figures and estimates in the main text, may be too wide. Figures C-3 and C-4 use the direct plug-in mean square error (MSE) bandwidth selector on only the stochastic side of the threshold and generate smaller bandwidths. Despite the different bandwidths, these point estimates, of 12.9 pp in Math and 57.1 pp in Portuguese, are close to the estimates in Table 9.

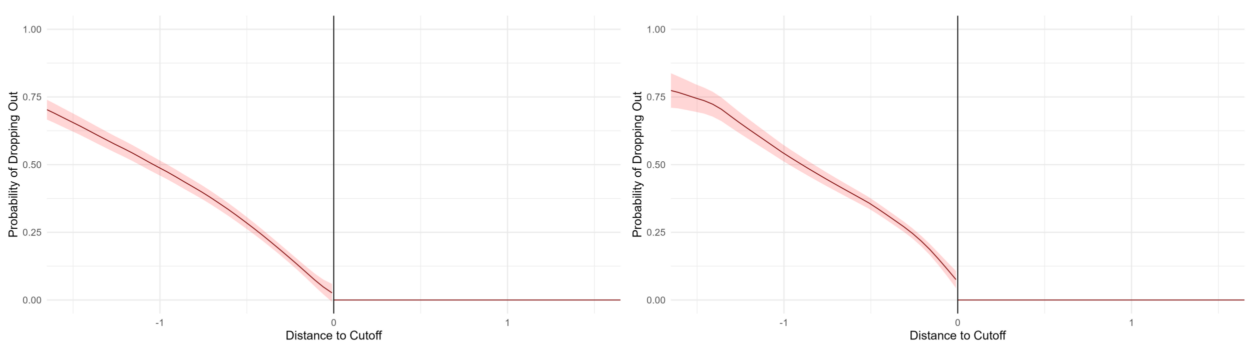
Figure 3: Histogram of Running Variables



(a) Math (b) Portuguese

This figure depicts histograms of the distance of the score from the passing threshold in math and Portuguese in a region around zero. Details regarding the construction of this variable are provided in section 7.3.

Figure 4: Probability of Dropout as a Function of the Running Variable



(a) Math (b) Portuguese

This figure depicts the relationship between the probability of dropout and the distance to the passing threshold in math in panel (a) and Portuguese in panel (b). Both nonparametric regressions are estimated using local linear regression with an Epanechnikov kernel and bandwidths selected following Calonico et al. (2017), 0.528 for math and 0.492 for Portuguese. The shaded regions represent 95% bias-corrected confidence intervals.

Table 9: The Effect of Retention on Secondary School Dropout

Policy	All	Boys	Girls
Status Quo (TT)	0.07	0.07	0.06
	(0.02)	(0.02)	(0.02)
10th grade retention (ATE)	0.04	0.06	0.03
	(0.01)	(0.01)	(0.01)
11th grade retention (ATE)	0.05	0.06	0.04
	(0.01)	(0.01)	(0.01)
12th grade retention (ATE)	0.10	0.10	0.09
	(0.01)	(0.01)	(0.01)
Mean: Overall	0.34	0.41	0.28
Mean: Retained Students	0.77	0.80	0.73
Effects at the Margin	Math	Portuguese	
RDD: Data	0.121 [-0.064 , 0.306]	0.525 [0.316 , 0.735]	
RDD: Model	0.116	0.427	

The table shows estimates of dropout treatment effects corresponding to four separate retention policies. The four policies are the status quo retention policy, and policies that retain all students in grades 10, 11, and 12 respectively. Standard errors, obtained from 200 parametric bootstrap replications, are shown in parentheses. The bottom of the table shows the regression discontinuity estimates of 12th grade retention in the data and as simulated by the model. Confidence intervals are 95% asymptotic confidence intervals.

in math. This pattern likely occurs because students at the margin of passing in Portuguese typically have weaker Portuguese scores than students at the margin of passing in mathematics, as the students studying in STEM fields (who take the mathematics exam) are on average stronger in both subjects.[19] The retention effects at the margin are larger than the average effects shown in Table 9 for two reasons. First, students much below the margin are more likely to drop out even in the absence of retention and, second, most retention in the 12th grade occurs in math, which has lower effects on dropout at the margin than Portuguese.

Table 9 showed that retention causes dropout and therefore reduces secondary school completion rates. Table B-12 in Appendix B shows that, under the status quo policy, 71%

[19]Students within two-tenths of the threshold of passing in math have a mean Portuguese score of 47.9, while the corresponding figure for students at the margin of passing Portuguese is 32.3.

of students, including 63% of boys and 77% of girls, complete secondary school within five years. An automatic grade promotion policy would instead have 84% of students complete school in five years. The effect of retention is therefore a 14 pp reduction in secondary school completion. The reduction is larger for boys (17 pp) than for girls (11 pp).[20]

7.5 Retention Impact Heterogeneity

Table 10 shows how retention's test score impacts vary with students' initial rankings in the ninth grade test score distribution, for the subgroup of students who experience retentions (see the table footnote). In both math and Portuguese, students who begin ninth grade with low test scores experience greater benefits from retention. For math, the effect is a 0.58 sd increase in test scores for students in the lowest tercile under the status quo policy. The effects on Portuguese scores are on average negative for all terciles, but less deleterious for students with initially lower ninth-grade test scores. Students with ninth-grade math scores in the top tercile also benefit from retention, with an average gain of 0.36 sd.[21]

Figures C-5 and C-6 in Appendix B plot the distributions of test score impacts for the four policies that we analyze. The standard deviation is large in all cases, which explains why retention causes test score reductions for substantial fractions of students. Despite the status quo policy increasing math scores by 0.46 sd on average, 18% of students experience a reduction in their 12th grade test scores. Similarly for Portuguese, despite retention reducing test scores on average by 0.17 sd, 42% of retained students benefit.

Table B-10 shows that the effect of retention on math test scores is similar by SES status, but retention causes worse treatment effects on Portuguese scores for higher SES students. Table B-11 shows that impacts vary by the student's age relative to the majority age at his/her grade level. Students who begin secondary school older than the modal age for their grade experience greater test score impacts in both math and Portuguese from retention.

[20]Retention reduces five-year secondary school completion by more than it raises dropout, because even after five years many retained students have neither completed nor dropped out in simulations.

[21]The percentage of top tercile students who are retained is 11.6% in math and 9.4% in Portuguese.

Table 10: Effects of Retention on Test Scores by 9th Grade Test Score Tercile

	Bottom Tercile		Middle Tercile		Top Tercile	
	Raw	S.D.	Raw	S.D.	Raw	S.D.
Math: 12th Grade						
Status Quo	13.72	0.58	10.44	0.44	8.37	0.36
	(1.39)	(0.06)	(0.72)	(0.03)	(0.97)	(0.04)
10th grade retention	0.15	0.01	-4.34	-0.18	-9.31	-0.40
	(1.66)	(0.07)	(1.00)	(0.04)	(1.00)	(0.04)
11th grade retention	3.36	0.14	-1.87	-0.08	-7.15	-0.30
	(0.94)	(0.04)	(0.71)	(0.03)	(0.88)	(0.04)
12th grade retention	6.13	0.26	0.85	0.04	-4.63	-0.20
	(0.5)	(0.02)	(0.56)	(0.02)	(0.78)	(0.03)
Portuguese: 12th Grade						
Status Quo	-0.32	-0.02	-5.34	-0.32	-9.85	-0.59
	(1.07)	(0.06)	(0.77)	(0.05)	(0.84)	(0.05)
10th grade retention	-5.54	-0.33	-10.03	-0.60	-15.2	-0.91
	(0.72)	(0.04)	(0.73)	(0.04)	(0.87)	(0.05)
11th grade retention	-5.97	-0.36	-10.58	-0.63	-15.87	-0.95
	(0.74)	(0.04)	(0.79)	(0.05)	(0.96)	(0.06)
12th grade retention	-7.67	-0.46	-12.01	-0.72	-17.25	-1.03
	(0.65)	(0.04)	(0.80)	(0.05)	(1.05)	(0.06)

The table presents the average effect of retention on test scores for students who graduate in exactly four years, broken down by their performance on the ninth grade exam for that subject.The standard deviations on the 12th grade math and Portuguese exams are 23.5 and 16.8 points.

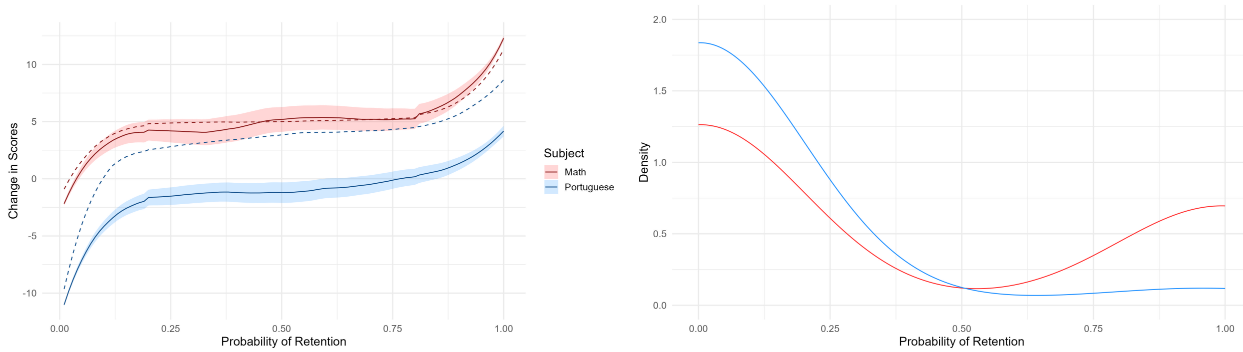
7.6 Selection into Retention

We next investigate whether schools are retaining the "right" students, that is, students who are most likely to benefit. To examine this, we simulate outcomes in a setting where everyone is retained in the 12th grade versus a scenario in which no one is retained in any grade and compare the resulting test scores at the time of graduation across the two scenarios. This exercise compares scores after four years of secondary school under the retention policy with scores after three years under the automatic promotion policy. We derive individual treatment impacts and nonparametrically regress these treatment impacts on the predicted probability of retention, or propensity score, for that student. Students with higher propensity scores are more likely to be retained.

Figure 5 displays the results of these nonparametric regressions in math and Portuguese. The figures are akin to marginal treatment effects curves, except in this case, the distribution of propensity scores plotted along the x-axis is not the unit uniform distribution (Heckman and Vytlacil 1999; 2005; 2007). The propensity score distributions are plotted in panel (b). The dotted line in panel (a) shows the non-parametric relationship between retention and test score impacts based on observed factors only (it constrains the estimated covariance matrix (5) to be diagonal), while the solid line and confidence intervals instead incorporate selection on both observed and unobserved factors by simulating from the full model including correlated shocks. The figure shows that observed factors, primarily test scores, generate an upward sloping relationship between retention probabilities and test score gains. Incorporating sorting on unobservables weakens this upward sloping relationship in Portuguese, but not in math. This is because the covariance matrix governing selection on unobservables, in Table 7, generates positive sorting on unobserved gains in math but negative sorting in Portuguese scores. Despite negative sorting on the basis of *unobservables* in Portuguese, the figure shows that sorting is positive overall in both subjects. The students with the greatest chance of being retained have the greatest test score benefits.

Figure 6 displays the estimated nonparametric regressions of the dropout treatment effect on the probability of being retained in math and Portuguese in the twelfth grade. As before, the model compares a world without retention to a world in which every student is retained in grade 12. The treatment effect is computed as the difference in indicator variables for whether the student dropped out. The figures show that, in both subjects, students who are more likely to be retained are more likely to drop out as a result of retention. These curves must be greater than zero everywhere, because a student who is promoted in the twelfth grade immediately completes secondary school. Nevertheless, there is meaningful variation in the probability of dropout both within and across subjects. Panel (b) shows that the causal effects of retention in Portuguese on dropout is about 50% for propensity scores above 0.2. This estimate is similar in magnitude to the RD estimates, which showed that students at the margin of being retained in Portuguese were 52.5 pp more likely to drop out after being retained.

Figure 5: Variation in 12th Grade Test Score Impacts by the Probability of Retention

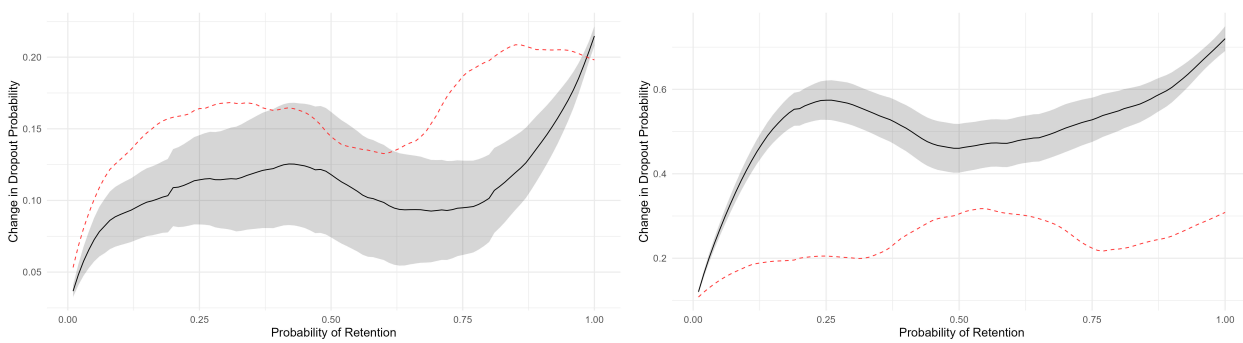


(a) 10th Grade

(b) Density

Panel (a) plots nonparametric regressions of the retention test score effects in math and Portuguese on the model-derived probability of retention in grade 12. The solid line and 95% confidence intervals are based on model simulations at the estimated parameters, while the dotted lines plot only sorting on observables. The regressions are estimated using local linear regression with an Epanechnikov kernel and a bandwidth of 0.20. Panel (b) depicts the density of twelfth grade retention probabilities by subject.

Figure 6: Variation in 12th Grade Dropout Impacts by the Probability of Retention



(a) Math

(b) Portuguese

The figures plot nonparametric regressions of the dropout effect of retention in math and Portuguese on the model-derived probability of retention in grade 12. The solid line and 95% confidence intervals in shaded grey are based on model simulations at the estimated parameters, while the dotted lines plot only sorting on observables. The regressions are estimated using local linear regression with an Epanechnikov kernel and a bandwidth of 0.20.

8 Towards an Optimal Retention Policy

We have seen that retention raises math test scores conditional on staying in school but also increases the probability of dropping out. An additional factor to consider is that retained students who stay in school will typically enter the labor market later and forego a year's salary or more. Any notion of an optimal retention policy must trade off the possible earnings benefits accruing from increased math ability with the costs incurred because of delayed labor market entry and potentially reduced educational attainment. Compulsory schooling laws can mitigate the costs to some extent if they are effective in preventing dropout. In this section, we consider grade retention's multiple effects on skill accumulation and educational attainment within a single framework and show how the optimal retention policy varies depending on the labor market returns to cognitive skills. Our analysis assumes that the compulsory schooling law, implemented in 2012/2013, is enforced, and that students may only drop out of school if they are age 18 or older.

Our cost-benefit calculations require predicting lifetime earnings under different retention policies, where retention can affect individuals' educational attainment, age of labor market entry, and cognitive skills. We obtain lifetime earnings profiles by estimating the following Mincer equation using Portugal's matched employer-employee data set, the *Quadros de Pessoal* (QP):

$$lnw_{i,t} = \alpha_0 + \alpha_1 S_i + (1 - HS_i)(\alpha_2 exp_{i,t} + \alpha_3 exp_{i,t}^2) + HS_i(\alpha_4 exp_{i,t} + \alpha_5 exp_{i,t}^2) + \varepsilon_{i,t} \;, \quad (9)$$

where S_i is years of schooling, $exp_{i,t}$ is experience, and the wage-experience profiles are allowed to differ based on whether the individual is a high school graduate or dropout. Because the QP does not measure cognitive skills in math or Portuguese, we link the MISI education data with the QP at the level of the concelho by computing concelho-level means of years of schooling and experience in the QP and of 12th grade math and Portuguese skills in MISI. We then estimate the returns to math and Portuguese skill in the following equation using the merged data set:

$$lnw_{j,t} = \widehat{\alpha_0} + \phi_1 K_j^M + \phi_2 K_j^P + \widehat{\alpha}_1 S_i + (1 - HS_j)(\widehat{\alpha}_2 exp_{j,t} + \widehat{\alpha}_3 exp_{j,t}^2) + \\ HS_j(\widehat{\alpha}_4 exp_{j,t} + \widehat{\alpha}_5 exp_{j,t}^2) + \varepsilon_{j,t} \ , \tag{10}$$

where j refers to the concelho, t denotes the year, and $\widehat{\alpha}_0, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_5$ are constrained to equal the values recovered from estimating equation (9). We combine data from two sources, estimating experience profiles and returns to schooling in the QP but estimating the returns to skills in the merged data set (which is aggregated to the municipality level), because there are not enough years in the merged data set to reliably estimate experience profiles.[22]

We use the estimated coefficients to simulate the wage for individual i in year t under a particular retention policy P as follows:

$$w_{it}(P) = \exp\Bigg(\widehat{\alpha}_0 + \widehat{\phi}_1 K_i^M + \widehat{\phi}_2 K_i^P + \widehat{\alpha}_1 S_i + (1 - HS_i)(\widehat{\alpha}_2 exp_{it} + \widehat{\alpha}_3 exp_{it}^2) + \\ HS_i(\widehat{\alpha}_4 exp_{it} + \widehat{\alpha}_5 exp_{it}^2) + \varepsilon_{it}\Bigg) \ . \tag{11}$$

All variables in equation (11) may be affected by the retention policy, P. Retention can influence years of schooling, students' terminal level of knowledge in math and Portuguese, and their on-the-job experience if the student is retained and enters the labor market a year later. We simulate each individual's profile of wages under retention policy P according to (11), convert wages to annual earnings, and calculate lifetime discounted earnings using the following expression:

$$Y_i(P) = \sum_{a=a_P}^{a=65} (\frac{1}{1+r})^{a-19} w_{i,t(a)}(P) \times 170 \times 14 \tag{12}$$

where a_P is the age of labor market entry, which may vary with the policy, thereby accounting for the foregone earnings due to delayed labor market entry.[23] The net benefit in terms of

[22] Table B-13 in Appendix B presents estimates from equations (9) and (10).

[23] We convert wages to an annual measure by multiplying them by 170 hours per month and 14 months of pay per year. Workers in Portugal receive both holiday pay and a Christmas bonus each worth one month of pay. Our cost-benefit calculation assume that individuals work every year after graduation.

lifetime earnings of the status quo policy, $\tilde{P}$, relative to a world with no retention is

$$\Delta = \frac{1}{N} \sum_{i=1}^{N} Y_i(\tilde{P}) - Y_i(0).$$

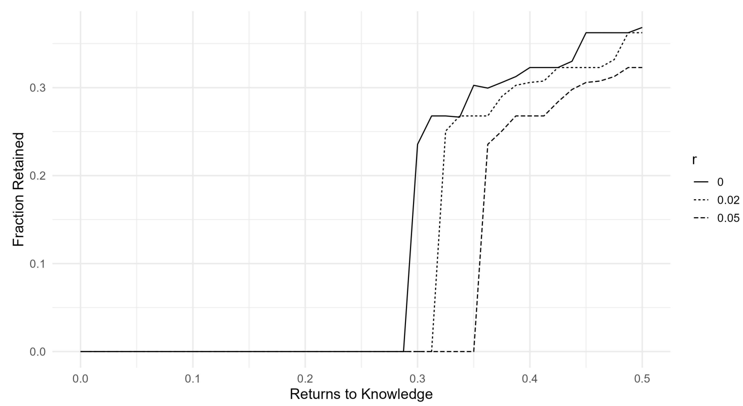
Δ corresponds to a policy-relevant treatment effect (PRTE), as defined in Heckman and Vytlacil (2005).[24] We solve for the policy $\tilde{P}$ that maximizes the PRTE by introducing a free parameter in the models' retention equation (3) that scales up or down retention rates and solving for the value of this variable that maximizes Δ. We find that the solution to this problem, given the estimated parameters in Table B-13, is to set the free parameter equal to negative infinity, which means that it is optimal not to retain students.

We next ask, what is the minimal return on math knowledge necessary to justify retaining a positive fraction of students? To answer this question, we fix the return to Portuguese knowledge at the value we estimated ($\widehat{\phi_2} = -0.00$) and vary ϕ_1 in equation (10). For each value of ϕ_1, we solve for the retention policy that maximizes the PRTE. Figure 7 plots the implied fraction of secondary school students who are retained in any grade under the optimal retention policy for values of ϕ_1 ranging between 0.00 and 0.50 and for interest rates equal to 0%, 2% and 5%. For low values of ϕ_1, it is optimal to not retain any students, but as ϕ_1 increases, the cognitive benefits of retention begin to outweigh the costs of reduced educational attainment and delayed labor market entry. As returns increase, it begins to be optimal to retain a positive fraction of students. For $r = 0$, a 28% return is necessary to justify retaining a positive fraction of students. For $r = 0.05$, the return to a one sd increase in math knowledge must be 35%. As the interest rate increases, the foregone earnings in the first year after leaving school take up a larger share of the present discounted value of lifetime income, making retention more costly. However, even for $r = 0$, the returns to math knowledge that could justify retaining a positive fraction of students are larger than what has been found in the literature.[25] The figure shows that high retention rates can only be

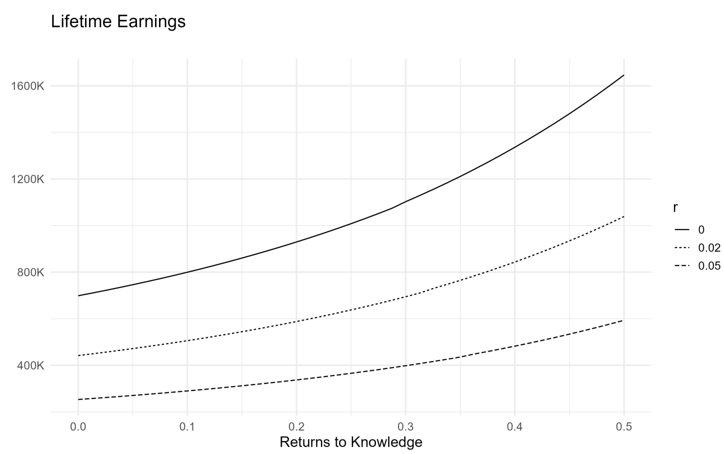
[24] Subsequent definitions of PRTEs, as in Mogstad et al. (2018), normalize Δ by the change in treatment probabilities across the two policy regimes. We work with the non-normalized PRTE, as we are interested in solving for the retention policy that maximizes the overall benefits for society.

[25] See, for example, Chetty et al. (2011), Dougherty (2003), Heckman et al. (2006), Cawley et al. (2001),

Figure 7: Optimal Retention Rates as a Function of Returns to Knowledge



(a) Retention rate



(b) Discounted lifetime income

The figures depicts the relationship between the returns to a standard deviation increase in math test scores along the x-axis and the fraction of students who are retained under the optimal retention policy in panel (a) and the expected discounted lifetime income under this policy in Euros (2013) in panel (b). The assumptions used to compute expected lifetime income are detailed in section 8.

justified from a cost-benefit perspective if the labor market richly rewards the skills that retention causally improves.

Our cost-benefit analysis is subject to a few caveats. First, the calculations assume

Murnane et al. (2000). A recent study by Watts (2020) uses a longitudinal UK data set, the National Child Development Study, to estimate associations between math and reading skills measured at age 16 and subsequent earnings at four ages between 33-50. The effects reported in the paper of a standard deviation increase in math test scores on yearly earnings range over the lifecycle from 3% to 11%.

that individuals are continuously employed. There is a small literature, including studies in economics, sociology and psychology, that analyzes whether retained students experience adverse labor market consequences in terms of a higher probability of unemployment and lower levels of job security. Baert and Picchio (2021) summarizes much of this literature and describes some studies finding evidence of adverse effects on employment.[26] Second, our calculations only consider private earnings returns to schooling and cognitive skills. The presence of substantial social returns could alter our conclusions about the optimal retention rate. Third, our cost-benefit analysis also did not account for possible positive impacts on future generations occurring through the intergenerational transmission of human capital. As seen in section 5, mother's and father's education levels are important determinants of youth's academic achievement. Allowing for intergenerational transmission of human capital would likely increase the benefits from retention. Finally, we consider the costs and benefits of retention for the student only. Taking into account the costs of educating retained students for an additional year will lower the societal benefits of retention and require still higher returns to knowledge to justify the existing policy.

9 Conclusions

In many OECD countries, it is common for students to experience grade retention, particularly in secondary school. Estimating the effects of grade retention on academic performance and on educational attainment poses significant challenges, because of the endogenous nature of the retention decision and the potential for retention to cause dropout. Students are not randomly assigned to be held back; rather, retention is influenced by a range of observable and unobservable factors, making it difficult to isolate its causal impact. Early studies on this topic often failed to account for this unobserved heterogeneity, limiting the validity of their conclusions.

[26] Baert and Picchio (2021) also present the results of an RCT that they carried out in Belgium that experimentally varied grade retention information on fictitious resumes. Their results showed that grade retention did not significantly affect positive employer call-back rates but made it less likely that individuals were called back for jobs with a large training component.

This paper develops and estimates an extended Roy model that simultaneously accounts for the endogeneity of the retention decision and potential dynamic selection bias caused by dropout. We apply the model to administrative data from Portugal to analyze secondary school retention's multiple effects on academic achievement and dropout. Estimates from the model show large positive average effects of retention on test scores, of 0.46 sd, for students retained in math who graduate in four years. Effects on Portuguese test scores are negative on average and smaller in magnitude, −0.17 sd. These average effects mask considerable heterogeneity, with 42% of students retained in Portuguese experiencing positive test score effects. We also find that retention discourages students from finishing secondary school, more so for students at the margin of retention than for the average retained student.

We conduct a cost-benefit analysis that compares lifetime earnings streams under different retention policies and find that, for typical values of the wage returns to knowledge that have been estimated in the literature, the current retention policy is not, on average, beneficial for students. The foregone earnings cost due to delayed labor market entry and reduced educational attainment for those induced to drop out exceeds the expected lifetime earnings gain accruing from greater knowledge. If retention is to be used as a policy tool for addressing problems related to low academic achievement, then it needs to be narrowly targeted at students who expect to benefit and to be combined with policies that prevent school dropout. We hope that the lessons derived from this study are valuable for policymakers, educators, researchers, and parents who seek to understand the broader impacts of grade retention on academic achievement, educational attainment, and earnings.

References

Agostinelli, F. and Wiswall, M. (2016). Estimating the Technology of Children's Skill Formation. Technical report, National Bureau of Economic Research.

Allen, C. S., Chen, Q., Willson, V. L., and Hughes, J. N. (2009). Quality of Research Design Moderates Effects of Grade Retention on Achievement: A Meta-Analytic, Multilevel Analysis. *Educational Evaluation and Policy Analysis*, 31(4):480–499.

Angrist, J. and Imbens, G. (1995). Identification and Estimation of Local Average Treatment Effects.

Åslund, O. and Rooth, D.-O. (2007). Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings. *The Economic Journal*, 117(518):422–448.

Baert, S. and Picchio, M. (2021). A Signal of (Train) Ability? Grade Repetition and Hiring Chances. *Journal of Economic Behavior & Organization*, 188:867–878.

Battistin, E. and Schizzerotto, A. (2019). Threat of Grade Retention, Remedial Education and Student Achievement: Evidence from Upper Secondary Schools in Italy. *Empirical Economics*, 56(2):651–678.

Bernal, R. (2008). The Effect of Maternal Employment and Child Care on Children's Cognitive Development. *International Economic Review*, 49(4):1173–1209.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). rdrobust: Software for Regression-Discontinuity Designs. *The Stata Journal*, 17(2):372–404.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, 131(2):633–686.

Carneiro, P., Hansen, K. T., and Heckman, J. J. (2003). Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College. *International Economic Review*, 44(2):361–422.

Cawley, J., Heckman, J., and Vytlacil, E. (2001). Three Observations on Wages and Measured Cognitive Ability. *Labour Economics*, 8(4):419–442.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Cockx, B., Picchio, M., and Baert, S. (2019). Modeling the Effects of Grade Retention in High School. *Journal of Applied Econometrics*, 34(3):403–424.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of the Economics of Education*, 1:697–812.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931.

De Groote, O. (2024). Dynamic Effort Choice in High School: Costs and Benefits of an Academic Track. *Journal of Labour Economics, forthcoming.*

Dougherty, C. (2003). Numeracy, Literacy and Earnings: Evidence from the National Longitudinal Survey of Youth. *Economics of Education Review*, 22(5):511–521.

Eide, E. R. and Showalter, M. H. (2001). The Effect of Grade Retention on Educational and Labor Market Outcomes. *Economics of Education Review*, 20(6):563–576.

Eren, O., Depew, B., and Barnes, S. (2017). Test-Based Promotion Policies, Dropping out, and Juvenile Crime. *Journal of Public Economics*, 153:9–31.

Ferreira, M., Golsteyn, B., and Parra-Cely, S. (2018). The Effect of Grade Retention on Secondary School Performance: Evidence from a Natural Experiment. *IZA Discussion Paper.*

Figlio, D. and Özek, U. (2020). An Extra Year to Learn English? Early Grade Retention and the Human Capital Development of English Learners. *Journal of Public Economics*, 186:104184.

Fruehwirth, J. C., Navarro, S., and Takahashi, Y. (2016). How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects. *Journal of Labor Economics*, 34(4):979–1021.

Gary-Bobo, R. J., Goussé, M., and Robin, J.-M. (2016). Grade Retention and Unobserved Heterogeneity. *Quantitative Economics*, 7(3):781–820.

Geweke, J. (1989). Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339.

Gourieroux, M., Gourieroux, C., Monfort, A., Monfort, D. A., et al. (1996). *Simulation-Based Econometric Methods*. Oxford University Press.

Hajivassiliou, V. A. and McFadden, D. L. (1998). The Method of Simulated Scores for the Estimation of LDV Models. *Econometrica*, pages 863–896.

Heckman, J. J. and Honore, B. E. (1990). The Empirical Content of the Roy Model. *Econometrica: Journal of the Econometric Society*, pages 1121–1149.

Heckman, J. J., LaLonde, R. J., and Smith, J. A. (1999). The Economics and Econometrics of Active Labor Market Programs. In *Handbook of Labor Economics*, volume 3, pages 1865–2097. Elsevier.

Heckman, J. J. and Navarro, S. (2007). Dynamic Discrete Choice and Dynamic Treatment Effects. *Journal of Econometrics*, 136(2):341–396.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Heckman, J. J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation 1. *Econometrica*, 73(3):669–738.

Heckman, J. J. and Vytlacil, E. J. (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences*, 96(8):4730–4734.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Handbook of Econometrics*, 6:4875–5143.

Holmes, C. T. and Matthews, K. M. (1984). The Effects of Nonpromotion on Elementary and Junior High School Pupils: A Meta-Analysis. *Review of Educational Research*, 54(2):225–236.

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

Jacob, B. A. and Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

Jimerson, S. R. (2001). Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century. *School Psychology Review*, 30(3):420–437.

Keane, M. P. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica: Journal of the Econometric Society*, pages 95–116.

Manacorda, M. (2012). The Cost of Grade Retention. *Review of Economics and Statistics*, 94(2):596–606.

Mogstad, M., Santos, A., and Torgovitsky, A. (2018). Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters. *Econometrica*, 86(5):1589–1619.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. H. (2000). How Important are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings? *Journal of Policy Analysis and Management*, 19(4):547–568.

Pereira, M. and Reis, H. (2014). Grade Retention During Basic Education in Portugal: Determinants and Impact on Student Achievement. *Economic Bulletin*, pages 61–82.

Portugal, P., Reis, H., Guimarães, P., and Cardoso, A. R. (2024). What Lies Behind the Returns to Shooling: the Role of Labor Market Sorting and Worker Heterogeneity. *Review of Economics and Statistics*, forthcoming.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2):135–146.

Saltiel, F. and Sarzosa, M. (2020). Grade Retention and Multidimensional Skill Formation in Young Children.

Schwerdt, G., West, M. R., and Winters, M. A. (2017). The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida. *Journal of Public Economics*, 152:154–169.

Watts, T. W. (2020). Academic Achievement and Economic Attainment: Reexamining Associations between Test Scores and Long-Run Earnings. *AERA Open*, 6(2).

Winters, M. A. and Greene, J. P. (2012). The Medium-Run Effects of Florida's Test-Based Promotion Policy. *Education Finance and Policy*, 7(3):305–330.

Zhong, J. (2024). Early Grade Retention Harms Adult Earnings. *Manuscript, Texas A & M University.*

A Moments Used in Estimation

This appendix describes the moments used in estimation. As discussed in the main text, the estimation is based on three types of moments for students with different schooling trajectories, as summarized by their history h_i, which was defined in section 5. The moments are formed from three types of simulated dependent variables: $\mathbb{1}_{h_i=h}$, $\mathbb{1}_{h_i=h}K_{i,t}$, and $\mathbb{1}_{h_i=h}(K_{i,t})^2$. That is, the moments depend on the proportion of students with each history, the mean knowledge levels (test scores) and the second moment of test scores for both math and Portuguese for students with different histories. Table B-2 lists the histories that form the basis of the SMM estimation for the STEM track students together with the percentage of students in the estimation sample that have each history.

Altogether, we target fifteen histories, which represent 89.1 percent of the STEM students in our estimation sample. For those histories in which 12th grade math and Portuguese test scores are observed (histories 1, 3, 4, 7, 9, 13, 14, and 15), we additionally target moments involving the 12th grade test scores.

Moments are formed by multiplying the three types of endogenous variables (histories, test scores, and squared test scores) by exogenous variables and summing over all individuals. The moments, m_i, for those individuals enrolled in the STEM track are as follows. The histories $h_1, \ldots, h_{15}$ correspond to the fifteen targeted STEM histories in Table B-2.

$$m_i = \begin{pmatrix} \mathbf{M_i^1} \\ \mathbf{M_i^2} \\ \mathbf{M_i^3} \end{pmatrix} \tag{13}$$

where

$$
\mathbf{M_i^1} = \begin{pmatrix}
\mathbb{1}_{h_i=h_1} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, \{Z_{i,t}^{D}\}_{t=1}^{3})' \\
\mathbb{1}_{h_i=h_2} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, I'_{i,1}, Z_{i,t}^{R}, \{Z_{i,t}^{D}\}_{t=1}^{2})' \\
\mathbb{1}_{h_i=h_3} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_4} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_5} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, \{Z_{i,t}^{D}\}_{t=1}^{3})' \\
\mathbb{1}_{h_i=h_6} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, \{Z_{i,t}^{D}\}_{t=1}^{3})' \\
\mathbb{1}_{h_i=h_7} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_8} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{2}, \{Z_{i,t}^{R}\}_{t=1}^{2}, \{Z_{i,t}^{D}\}_{t=1}^{3})' \\
\mathbb{1}_{h_i=h_9} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_{10}} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_{11}} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_{12}} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, I'_{i,1}, Z_{i,t}^{R}, \{Z_{i,t}^{D}\}_{t=1}^{1})' \\
\mathbb{1}_{h_i=h_{13}} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_{14}} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, \{Z_{i,t}^{D}\}_{t=1}^{4})' \\
\mathbb{1}_{h_i=h_{15}} \cdot (1, K_{i,0}^{9,M}, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, \{Z_{i,t}^{D}\}_{t=1}^{4})'
\end{pmatrix},
$$

The moments in $\mathbf{M^1}$ represent simulated analogs of covariances between the variables in the retention and dropout equations and the history indicators, which identify the parameters of the dropout and retention equations. These moments correspond to simulated versions of the first order conditions of the likelihood contributions of each of the 15 paths, where the probability density of each path has been replaced by the simulated indicator, $\mathbb{1}_{h_i=h}$. The moments in $\mathbf{M^2}$ are simulated analogs of covariances between the covariates in the value-added equations and the 12th grade math and Portuguese test scores for both retained and non-retained students. These moments identify the parameters in the value-added equations. The moments in $\mathbf{M^3}$ are squared test scores for students with each history in which test scores are observable. These moments identify the variances of the error terms in the value-added equations.

$$
\mathbf{M_i^2} = \begin{pmatrix}
\mathbb{1}_{h_i=h_1} K_{i,3}^{12,M} \cdot (1, K_{i,0}^{9,M}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_1} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_3} K_{i,3}^{12,M} \cdot (1, K_{i,0}^{9,M}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_3} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_3} K_{i,4}^{12,M} \cdot (1, K_{i,0}^{9,M}, I'_{i,4})' \\
\mathbb{1}_{h_i=h_4} K_{i,3}^{12,M} \cdot (1, K_{i,0}^{9,M}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_4} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_4} K_{i,4}^{12,M} \cdot (1, K_{i,0}^{9,M}, I'_{i,4})' \\
\mathbb{1}_{h_i=h_7} K_{i,3}^{12,M} \cdot (1, K_{i,0}^{9,M}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_7} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_9} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^4)' \\
\mathbb{1}_{h_i=h_{13}} K_{i,3}^{12,M} \cdot (1, K_{i,0}^{9,M}, \{I'_{i,t}\}_{t=1}^4)' \\
\mathbb{1}_{h_i=h_{13}} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^4)' \\
\mathbb{1}_{h_i=h_{14}} K_{i,3}^{12,M} \cdot (1, K_{i,0}^{9,M}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_{14}} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^3)' \\
\mathbb{1}_{h_i=h_{14}} K_{i,4}^{12,M} \cdot (1, K_{i,0}^{9,M}, I'_{i,4})' \\
\mathbb{1}_{h_i=h_{14}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, I'_{i,4})'
\end{pmatrix}, \qquad
\mathbf{M_i^3} = \begin{pmatrix}
\mathbb{1}_{h_i=h_1} K_{i,3}^{12,M^2} \\
\mathbb{1}_{h_i=h_1} K_{i,3}^{12,P^2} \\
\mathbb{1}_{h_i=h_3} K_{i,3}^{12,M^2} \\
\mathbb{1}_{h_i=h_3} K_{i,3}^{12,P^2} \\
\mathbb{1}_{h_i=h_3} K_{i,4}^{12,M^2} \\
\mathbb{1}_{h_i=h_4} K_{i,3}^{12,M^2} \\
\mathbb{1}_{h_i=h_4} K_{i,3}^{12,P^2} \\
\mathbb{1}_{h_i=h_4} K_{i,4}^{12,M^2} \\
\mathbb{1}_{h_i=h_7} K_{i,3}^{12,M^2} \\
\mathbb{1}_{h_i=h_7} K_{i,3}^{12,P^2} \\
\mathbb{1}_{h_i=h_9} K_{i,4}^{12,P^2} \\
\mathbb{1}_{h_i=h_{13}} K_{i,3}^{12,M^2} \\
\mathbb{1}_{h_i=h_{13}} K_{i,3}^{12,P^2} \\
\mathbb{1}_{h_i=h_{14}} K_{i,3}^{12,M^2} \\
\mathbb{1}_{h_i=h_{14}} K_{i,3}^{12,P^2} \\
\mathbb{1}_{h_i=h_{14}} K_{i,4}^{12,M^2} \\
\mathbb{1}_{h_i=h_{14}} K_{i,4}^{12,P^2}
\end{pmatrix}.
$$

Our vector of moments used in estimation also includes moments for students in the non-STEM track. For non-STEM students, the test score moments only pertain to Portuguese scores, because these students do not take math exams. Table B-3 lists the histories that we target for the students enrolled in the non-STEM track together with the percentage of the students in the estimation sample that have each history.

The nine histories in table B-3 represent 98.7 percent of the non-STEM students in our estimation sample. For these non-STEM students, we include 12th grade Portuguese test score moments, $\mathbb{1}_{(h_i=h)} K_{i,t}^P$ and $\mathbb{1}_{(h_i=h)} (K_{i,t}^P)^2$ for those histories which permit a 12th grade Portuguese test score to be observed (all histories except history 8, in which test scores are only observed after the 5th year of secondary school enrollment and thus beyond the reach of our data, which covers ninth grade plus test scores measured up to five years later).

Moments for these non-STEM students are formed in the same way as for STEM students, by multiplying the endogenous variables (history indicators, test scores, and squared test scores) by the covariates in the value-added, retention, and dropout equations and summing over all individuals. These moments are given below, where the histories $h_1^{NS}, \ldots h_9^{NS}$ correspond to the nine targeted non-STEM histories shown in Table B-3.

$$m_i^{NS} = \begin{pmatrix} \mathbf{M_i^{1,NS}} \\ \mathbf{M_i^{2,NS}} \\ \mathbf{M_i^{3,NS}} \end{pmatrix}$$

where

$$\mathbf{M_i^{1,NS}} = \begin{pmatrix} \mathbb{1}_{h_i=h_1^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3})' \\ \mathbb{1}_{h_i=h_2^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, Z_{i,4}^{D})' \\ \mathbb{1}_{h_i=h_3^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3}, Z_{i,4}^{D})' \\ \mathbb{1}_{h_i=h_4^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4} Z_{i,4}^{D})' \\ \mathbb{1}_{h_i=h_5^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, Z_{i,4}^{D})' \\ \mathbb{1}_{h_i=h_6^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4})' \\ \mathbb{1}_{h_i=h_7^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, Z_{i,4}^{D})' \\ \mathbb{1}_{h_i=h_8^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3}, \{Z_{i,t}^{R}\}_{t=1}^{3})' \\ \mathbb{1}_{h_i=h_9^{NS}} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4}, \{Z_{i,t}^{R}\}_{t=1}^{4}, Z_{i,4}^{D})' \end{pmatrix},$$

$$\mathbf{M_i^{2,NS}} = \begin{pmatrix} \mathbb{1}_{h_i=h_1^{NS}} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3})' \\ \mathbb{1}_{h_i=h_2^{NS}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4})' \\ \mathbb{1}_{h_i=h_3^{NS}} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3})' \\ \mathbb{1}_{h_i=h_4^{NS}} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3})' \\ \mathbb{1}_{h_i=h_4^{NS}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, I'_{i,4})' \\ \mathbb{1}_{h_i=h_5^{NS}} K_{i,3}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{3})' \\ \mathbb{1}_{h_i=h_5^{NS}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, I'_{i,4})' \\ \mathbb{1}_{h_i=h_6^{NS}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4})' \\ \mathbb{1}_{h_i=h_7^{NS}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4})' \\ \mathbb{1}_{h_i=h_9^{NS}} K_{i,4}^{12,P} \cdot (1, K_{i,0}^{9,P}, \{I'_{i,t}\}_{t=1}^{4})' \end{pmatrix}, \quad \mathbf{M_i^{3,NS}} = \begin{pmatrix} \mathbb{1}_{h_i=h_1^{NS}} K_{i,3}^{12,P^2} \\ \mathbb{1}_{h_i=h_2^{NS}} K_{i,4}^{12,P^2} \\ \mathbb{1}_{h_i=h_3^{NS}} K_{i,3}^{12,P^2} \\ \mathbb{1}_{h_i=h_4^{NS}} K_{i,3}^{12,P^2} \\ \mathbb{1}_{h_i=h_4^{NS}} K_{i,4}^{12,P^2} \\ \mathbb{1}_{h_i=h_5^{NS}} K_{i,3}^{12,P^2} \\ \mathbb{1}_{h_i=h_5^{NS}} K_{i,4}^{12,P^2} \\ \mathbb{1}_{h_i=h_6^{NS}} K_{i,4}^{12,P^2} \\ \mathbb{1}_{h_i=h_7^{NS}} K_{i,4}^{12,P^2} \\ \mathbb{1}_{h_i=h_9^{NS}} K_{i,4}^{12,P^2} \end{pmatrix}.$$

As before, the moments in $\mathbf{M^{1,NS}}$ represent simulated analogs of covariances between the covariates in the retention and dropout equations and the history indicators, the moments in $\mathbf{M^{2,NS}}$ are simulated analogs of covariances between the covariates in the value-added equations and the endogenous twelfth grade Portuguese test scores for both retained and non-retained students, and the moments in $\mathbf{M^{3,NS}}$ are squared test scores for students with each history in which test scores are observable. The moments in $\mathbf{M^{1,NS}}$ correspond to simulated versions of the first order conditions of the likelihood contributions of each of the 9 non-STEM paths, where the probability density of each path has been replaced by the simulated indicator, $\mathbb{1}_{h_i=h^{NS}}$.

B Additional Tables

Table B-1: Comparing the Full and Estimation Samples

	Estimation			Full		
	Mean	SD	Fraction Overall	Mean	SD	Fraction Overall
Math Score (9th grade)	62.4	20.4	0.68	61.2	20.7	0.97
Portuguese Score (9th grade)	59.9	14.8	0.68	59.3	14.7	0.97
Low SES	0.33	0.47	0.68	0.34	0.47	1
Mother < Basic Educ	0.45	0.50	0.68	0.45	0.50	1
Mother Basic Educ	0.18	0.38	0.68	0.18	0.39	1
Mother Sec. Educ	0.19	0.39	0.68	0.19	0.39	1
Mother > Sec. Educ	0.18	0.39	0.68	0.18	0.38	1

The table shows the means and standard deviations for several variables that are rarely missing in both the estimation and the full samples. The columns labeled Fraction Overall indicate the fraction of the full sample that is used to compute the means and standard deviations for Estimation and Full samples. Ninth grade test scores are only present in 97% of observations in the sample.

Table B-2: Model Fit, STEM Histories

History	Data	Crude Simulator	Smooth Simulator
10-11-12X10-11-12	0.379	0.410	0.408
10-10dX10-10d	0.066	0.074	0.075
10-11-12-12-12X10-11-12	0.039	0.030	0.029
10-11-12-12X10-11-12	0.036	0.026	0.026
10-11-12dX10-11-12d	0.016	0.013	0.013
10-10-10dX10-10-10d	0.019	0.024	0.024
10-11-12-12dX10-11-12	0.023	0.018	0.018
10-11-11dX10-11-11d	0.016	0.019	0.019
10-10-11-12X10-10-11-12	0.025	0.012	0.013
10-10-11-11dX10-10-11-11d	0.016	0.007	0.007
10-11-11-11dX10-11-11-11d	0.012	0.002	0.003
10-11dX10-11d	0.012	0.008	0.009
10-11-11-12X10-11-11-12	0.009	0.003	0.004
10-11-12-12X10-11-12-12	0.001	0.001	0.001
10-11-12-12-12dX10-11-12-12-12d	0.002	0.001	0.001

The table shows the in-sample fit of the model for a subset of targeted moments: the proportion of STEM students with each history. The smooth simulator uses a logit smoother with a scale parameter of $\lambda = 0.1$.

Table B-3: Model Fit, non-STEM Histories

History	Data	Crude Simulator	Smooth Simulator
10-11-12	0.187	0.216	0.215
10-10-11-12	0.026	0.009	0.010
10-11-12-12d	0.008	0.009	0.009
10-11-12-12-12d	0.006	0.003	0.003
10-11-12-12	0.005	0.002	0.002
10-11-11-12	0.004	0.001	0.001
10-10-11-12-12d	0.004	0.002	0.002
10-10-11-11-12	0.002	0.001	0.001
10-11-11-12-12d	0.001	0.001	0.001

The table shows the in-sample fit of the model for a subset of targeted moments: the proportion of non-STEM students with each history. The smooth simulator uses a logit smoother with a scale parameter of $\lambda = 0.1$.

Table B-4: Regressions of Course Mark Residuals on Retention Exclusion Restriction

	Math	Portuguese
	(1)	(2)
Historical Retention Rate	−1.872	−2.141
	(0.927)	(0.983)
Constant	0.353	0.432
	(0.211)	(0.223)
Observations	252	252
R^2	0.016	0.018

The table shows estimates from regressions of course mark residuals on the retention exclusion restriction, Z^R. To produce the dependent variable, marks in each course in the 12th grade are regressed on a quadratic in test scores and the resulting residuals are averaged by concelho. These concelho-specific mean residuals are then regressed on the retention rate in that concelho three years earlier.

Table B-5: Course Marks and the Decision to Drop Out

	Drop Out (Probit)			
	(1)	(2)	(3)	(4)
Mark: Math (0-20)	-0.018	-0.015	-0.027	Categorical
	(0.036)	(0.036)	(0.036)	
Mark: PT (0-20)	-0.087	-0.011	-0.015	Categorical
	(0.015)	(0.018)	(0.018)	
Exam: Math (0-100)		-0.032	-0.029	-0.029
		(0.003)	(0.003)	(0.003)
Exam: PT (0-100)		-0.010	-0.007	-0.007
		(0.002)	(0.002)	(0.002)
Observations	3,565	3,565	3,565	3,565
Controls	No	No	Yes	Yes
Joint F-Test p-value	0.00	0.751	0.518	0.833

The table shows estimates from probit models predicting drop out based on course marks and exam scores in math and Portuguese. The regressions are estimated on the subsample of individuals who are retained in the 12th grade in either math or Portuguese. Controls include 9th grade test scores and all variables in Table 1. The fourth column includes indicator variables for each value that marks can take. The bottom row presents the p-value from a joint hypothesis test that marks do not predict dropout.

Table B-6: Estimated Value-added Equation Parameters, Alternative Specification

	Math				*Portuguese*			
	First time		Repeating		First time		Repeating	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Lagged Score, $K_{i,t-1}$	0.979	0.004	0.599	0.02	0.835	0.003	0.575	0.033
Constant	-9.379	0.593	13.24	2.623	2.867	0.485	-2.684	7.60
Relative Age at Gr 10	-3.774	0.193	-0.12	0.589	-1.771	0.076	1.301	0.78
Female Teacher	-0.788	0.16	1.643	0.811	0.238	0.15	0.376	2.287
Teacher Age	0.352	0.09	0.155	0.482	0.549	0.085	2.776	1.425
Class size	-0.015	0.152	0.007	0.568	0.024	0.088	-0.015	1.206
Low SES	-0.963	0.133	-0.454	0.588	-0.156	0.082	3.64	1.275
Mother Basic Educ	0.137	0.145	0.885	0.857	-0.089	0.087	-0.631	1.693
Mother Sec. Educ	0.59	0.139	0.331	0.829	0.126	0.088	-3.427	1.864
Mother > Sec. Educ	2.366	0.139	1.917	1.047	1.303	0.093	-2.449	2.319
Semiurban	0.256	0.139	-1.08	0.878	0.227	0.089	0.336	1.869
Urban	-0.108	0.141	0.656	0.857	-0.056	0.088	0.944	1.702
Female	1.141	0.097	1.835	0.627	0.71	0.064	2.581	1.291

The table presents parameter estimates for the four value-added equations. The omitted education category is less than basic. Class size is measured in tens of students, and teacher age is measured in tens of years. This specification uses the contemporaneous retention rate at other schools within the student's concelho as the exclusion restriction.

Table B-7: Dropout Equation Parameters, Alternative Specification

	Estimate	SE
Intercept - 10th grade	1.078	0.104
Intercept - 11th grade	1.137	0.105
Intercept - 12th grade	0.839	0.095
Math	-3.051	0.081
Portuguese	-2.198	0.114
Local Income: Dropout - Graduate	0.057	0.015
Relative Age at Gr 10	0.379	0.018
Yrs Retained in Sec. School	-0.001	0.037
Class size	0.007	0.024
Low SES	0.057	0.021
Mother Basic Educ	-0.075	0.026
Mother Sec. Educ	-0.149	0.028
Mother > Sec. Educ	-0.134	0.036
Semiurban	0.059	0.029
Urban	0.124	0.029
Female	-0.047	0.022
Non-STEM	-1.196	0.073

The table presents parameter estimates for the dropout equations. The omitted education category is less than basic. The coefficients and standard errors on math and Portuguese knowledge have been scaled so that they represent the effects of a 100-point increase in these scores. Class size is measured in tens of students. Income is measured in 100s of Euros per month. This specification uses the contemporaneous retention rate at other schools within the student's concelho as the exclusion restriction.

Table B-8: Retention Equation Parameters, Alternative Specification

	Estimate	SE
Intercept - 10th grade	3.026	0.118
Intercept - 11th grade	1.789	0.034
Intercept - 12th grade, Math	4.167	0.187
Intercept - 12th grade, Portuguese	0.928	0.126
Slope - Math, 10/11th grades	-4.406	0.129
Slope - Math, 12th grade	-16.47	0.531
Slope - Portuguese, 10/11th grades	-3.927	0.146
Slope - Portuguese, 12th grade	-8.460	0.199
Retention Rate (other schools)	1.409	0.157
Relative Age at Gr 10	0.247	0.021
Class size	0.102	0.024
Low SES	-0.075	0.021
Mother Basic Educ	-0.039	0.026
Mother Sec. Educ	-0.062	0.027
Mother > Sec. Educ	0.002	0.033
Semiurban	-0.037	0.028
Urban	0.102	0.027
Female	-0.019	0.022
Non-STEM	-4.433	0.093

The table presents parameter estimates for the retention equations. Retention is grade-specific in grades 10 and 11 but grade-subject-specific in grade 12. Retention in grade 12 in Math depends on math scores but not Portuguese scores and vice versa for retention in grade 12 in Portuguese. The omitted education category is less than basic. The coefficients and standard errors on math and Portuguese knowledge have been scaled so that they represent the effects of a 100-point increase in these scores. This specification uses the contemporaneous retention rate at other schools within the student's concelho as the exclusion restriction.

Table B-9: Estimated Covariance Matrix, Alternative Specification

$\varepsilon^M(0)$	$\varepsilon^M(1)$	$\varepsilon^P(0)$	$\varepsilon^P(1)$	Dropout Shock	Retention Shock
115.822	0	49.495	0	0.777	0.711
(1.595)	–	(1.444)	–	(0.359)	(0.351)
	144.519	0	-75.598	10.057	5.09
	(5.474)	–	(13.658)	(0.897)	(0.628)
		72.017	0	1.299	4.375
		(0.739)	–	(0.306)	(0.271)
			44.087	-6.303	-3.735
			(11.689)	(0.922)	(0.781)
				1	0.592
					(0.023)
					1

The table presents parameter estimates for the covariance matrix of contemporaneous shocks in (5). Standard errors are in parentheses. This specification uses the contemporaneous retention rate at other schools within the student's concelho as the exclusion restriction for retention.

Table B-10: Retention Impacts by Income

	Math		*Portuguese*		
Low Income	Raw	S.D.	Raw	S.D.	Prop. with score
Status Quo	11.63	0.49	-1.32	-0.08	0.08
	(0.82)	(0.04)	(1.04)	(0.06)	(0.01)
Retain 10	-5.33	-0.23	-8.36	-0.50	0.74
	(1.1)	(0.05)	(0.82)	(0.05)	(0.02)
Retain 11	-2.48	-0.11	-8.68	-0.52	0.73
	(0.74)	(0.03)	(0.87)	(0.05)	(0.02)
Retain 12	0.31	0.01	-10.00	-0.60	0.66
	(0.56)	(0.02)	(0.87)	(0.05)	(0.01)
High Income					
Status Quo	10.53	0.45	-3.84	-0.23	0.07
	(0.67)	(0.03)	(1.17)	(0.07)	(0.01)
Retain 10	-6.46	-0.27	-11.83	-0.71	0.83
	(0.98)	(0.04)	(0.77)	(0.05)	(0.02)
Retain 11	-4.13	-0.18	-12.47	-0.74	0.82
	(0.78)	(0.03)	(0.86)	(0.05)	(0.02)
Retain 12	-1.65	-0.07	-14.18	-0.85	0.79
	(0.69)	(0.03)	(0.92)	(0.05)	(0.01)

The table shows estimates of test score treatment effects corresponding to four separate retention policies disaggregated by household income for students who graduate from secondary school in four years. Low income means that a student qualifies for a public income subsidy based on their parent's income. The standard deviations on the 12th grade math and Portuguese exams are 23.6 and 17.0 points. The column labeled proportion indicates the proportion in the simulation who take an exam in the 12th grade. Standard errors, obtained from 200 parametric bootstrap replications, are shown in parentheses.

Table B-11: Retention Impacts by Age Relative to Peers

	Math		*Portuguese*		
At Grade Level	Raw	S.D.	Raw	S.D.	Grad in 4 Years
Status Quo (TT)	10.73	0.46	-3.91	-0.23	0.07
	(1.29)	(0.05)	(0.89)	(0.05)	
Retain 10	-6.23	-0.27	-11.46	-0.68	0.86
	(1.99)	(0.09)	(0.74)	(0.04)	
Retain 11	-3.90	-0.17	-12.17	-0.73	0.86
	(1.31)	(0.06)	(0.79)	(0.05)	
Retain 12	-1.40	-0.06	-13.72	-0.82	0.83
	(0.68)	(0.03)	(0.70)	(0.04)	
One Year Older					
Status Quo	12.39	0.53	-0.10	-0.01	0.07
	(0.65)	(0.03)	(1.12)	(0.07)	
Retain 10	-5.02	-0.21	-6.47	-0.39	0.60
	(0.94)	(0.04)	(0.77)	(0.05)	
Retain 11	-0.77	-0.03	-6.08	-0.36	0.57
	(0.74)	(0.03)	(0.85)	(0.05)	
Retain 12	2.82	0.12	-7.01	-0.42	0.47
	(0.65)	(0.03)	(0.88)	(0.05)	
Two or More Years Older					
Status Quo	13.58	0.58	2.98	0.18	0.05
	(1.83)	(0.08)	(0.97)	(0.06)	
Retain 10	-6.32	-0.27	-3.55	-0.21	0.35
	(3.43)	(0.15)	(0.89)	(0.05)	
Retain 11	0.06	0.00	-1.98	-0.12	0.34
	(2.42)	(0.1)	(1.03)	(0.06)	
Retain 12	5.18	0.22	-2.42	-0.14	0.20
	(1.16)	(0.05)	(1.12)	(0.07)	

The table shows estimates of test score treatment effects corresponding to four separate retention policies disaggregated by the age at which the student enters secondary school. Students who enter secondary school one or more year above grade level have typically been retained prior to entering secondary school. The standard deviations on the 12th grade math and Portuguese exams are 23.5 and 16.8 points. Standard errors, obtained from 200 parametric bootstrap replications, are shown in parentheses.

Table B-12: Effect of Retention on Secondary School Completion

	All	*Boys*	*Girls*
Status Quo	0.71	0.63	0.77
	(0.01)	(0.01)	(0.01)
No Retention	0.84	0.81	0.87
	(0.01)	(0.01)	(0.01)
Effect of Retention	-0.14	-0.17	-0.11
	(0.01)	(0.01)	(0.01)

The table presents estimates of the effect of the status quo grade retention policy on the fraction of secondary school students who complete it within five years. Standard errors, obtained from 200 parametric bootstrap replications, are shown in parentheses.

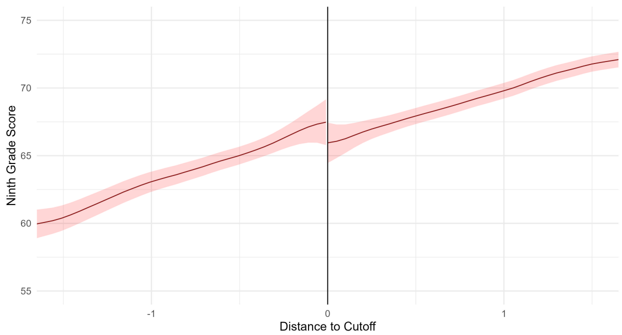
Table B-13: Estimated Parameters, Log Wage Equation

	Coefficient	S.E.	Data Set
$K_{i,12}^M$	0.142	(0.015)	MISI-QP concelho-level merge
$K_{i,12}^P$	-0.00	(0.017)	MISI-QP concelho-level merge
Education (years)	0.037	(0.000)	QP
Experience: $<$ secondary school	.032	(0.000)	QP
Experience^2: $<$ secondary school	-.00	(0.000)	QP
Experience: $\geq$ secondary school	.043	(0.000)	QP
Experience^2: $\geq$ secondary school	-.00	(0.000)	QP
Constant	0.899	(0.002)	QP
Observations	3,101		
RMSE	0.097		
R^2	0.607		

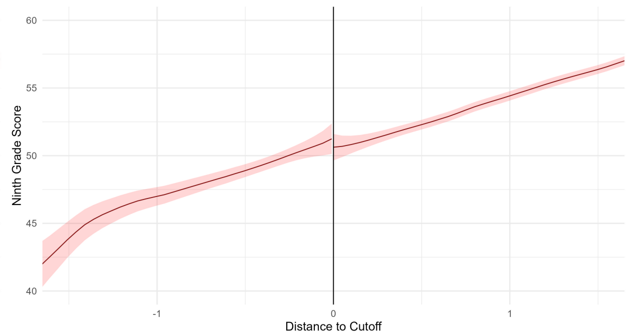
The table presents estimates of the Mincer equation relating schooling, experience, and skills in math and Portuguese to the logarithm of hourly wages. The regression allows for separate experience profiles for high school dropouts and high school graduates. The estimates come from two data sets: returns to schooling and experience profiles are estimated using the individual data in the Quadros de Pessoal (QP). Then, concelho level means of these variables are computed in the QP and merged to the MISI educational data at the level of the concelho for the cohort entering the labor market in 2008. The coefficients on $K_{i,12}^M$ and $K_{i,12}^P$ are then estimated in this merged data set using a regression that constrains the returns to schooling and experience profiles at their values recovered from the QP. This approach was employed because we do not observe workers long enough in the MISI-QP merge to reliably estimate experience profiles.

C Additional Figures

Figure C-1: Covariate Balanced Around the Cutoff



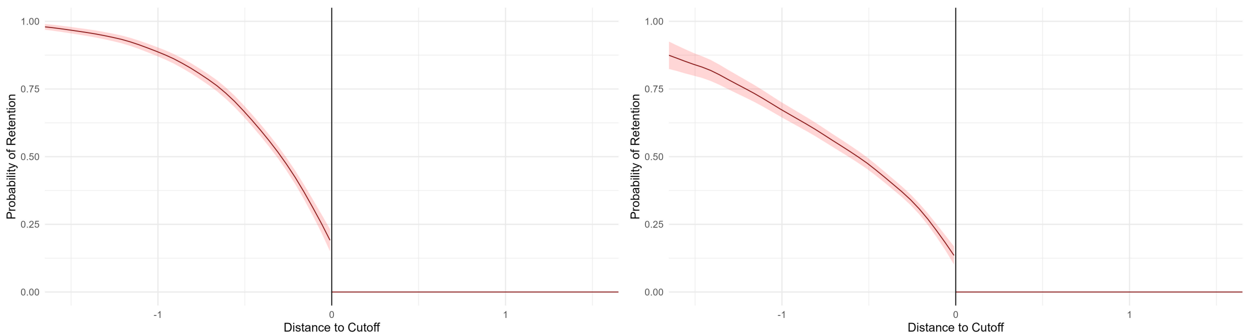
(a) Math



(b) Portuguese

The figure examines covariate balance around the passing threshold in math and Portuguese. Panel (a) plots ninth grade math scores along the y-axis against the score used to determine retention in the 12th grade in math. Panel (b) plots ninth grade Portuguese scores along the y-axis against the score used to determine retention in the 12th grade in Portuguese. Both nonparametric regressions are estimated using local linear regression with an Epanechnikov kernel and the same bandwidths as in the main text, 0.528 for math and 0.492 for Portuguese. The shaded regions represent 95% bias-corrected confidence intervals.

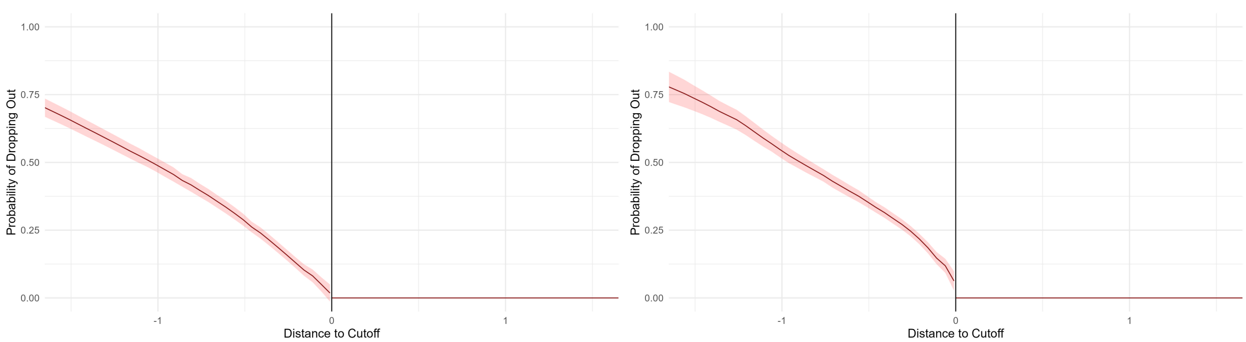
Figure C-2: Probability of Retention as a Function of the Running Variable



(a) Math (b) Portuguese

This figure depicts the relationship between the probability of retention and the distance of the score to the cutoff in math in panel (a) and Portuguese in panel (b). Both nonparametric regressions are estimated using local linear regression with an Epanechnikov kernel and the same bandwidths as in the main text, 0.528 for math and 0.492 for Portuguese. The shaded regions represent 95% bias-corrected confidence intervals.

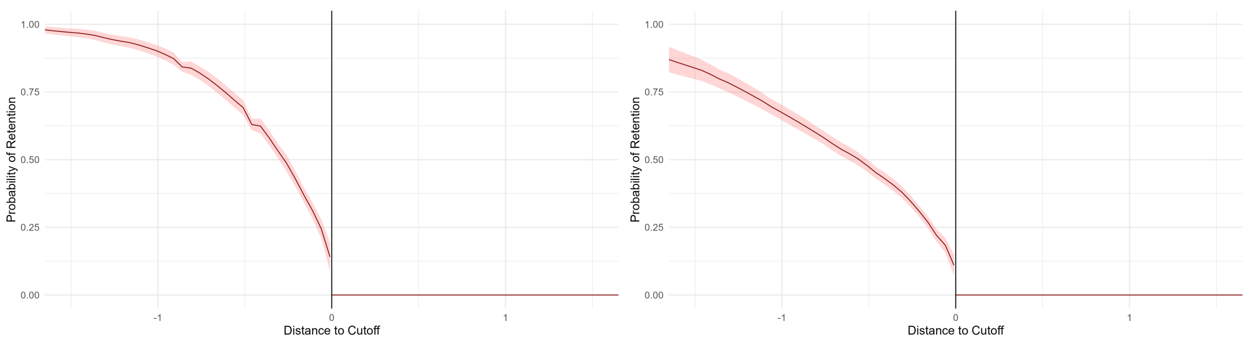
Figure C-3: Second Stage Regressions: Alternative Bandwidth



(a) Math (b) Portuguese

This figure depicts the relationship between the probability of dropout and the distance of the score to the passing threshold in math in panel (a) and Portuguese in panel (b) using a different bandwidth from Figure 4. The bandwidth for each subject is chosen using the direct plug-in MSE-optimal bandwidth selector on the left hand side of the passing threshold (where dropout is variable). This results in a separate bandwidth at each evaluation point. The bandwidth at immediately to the left of the threshold is 0.633 for math and 0.361 for Portuguese. Both panels estimate local linear regressions with an Epanechnikov kernel.

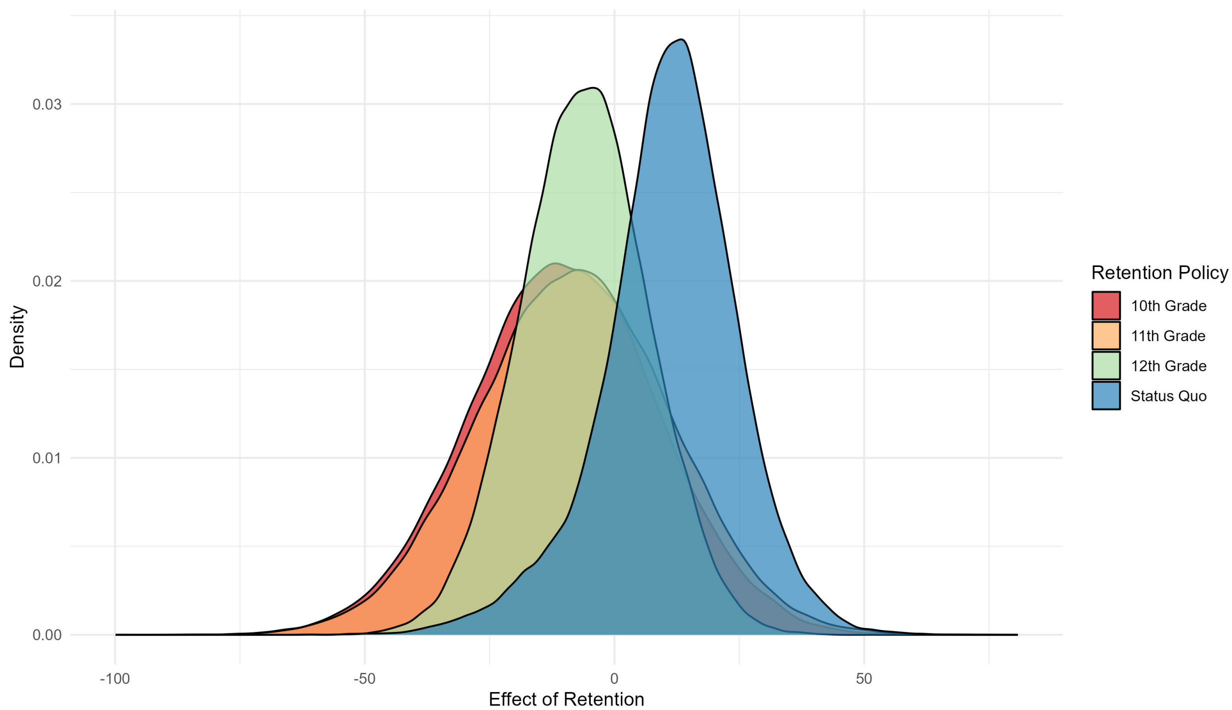
Figure C-4: First Stage Regressions: Alternative Bandwidth



(a) Math (b) Portuguese

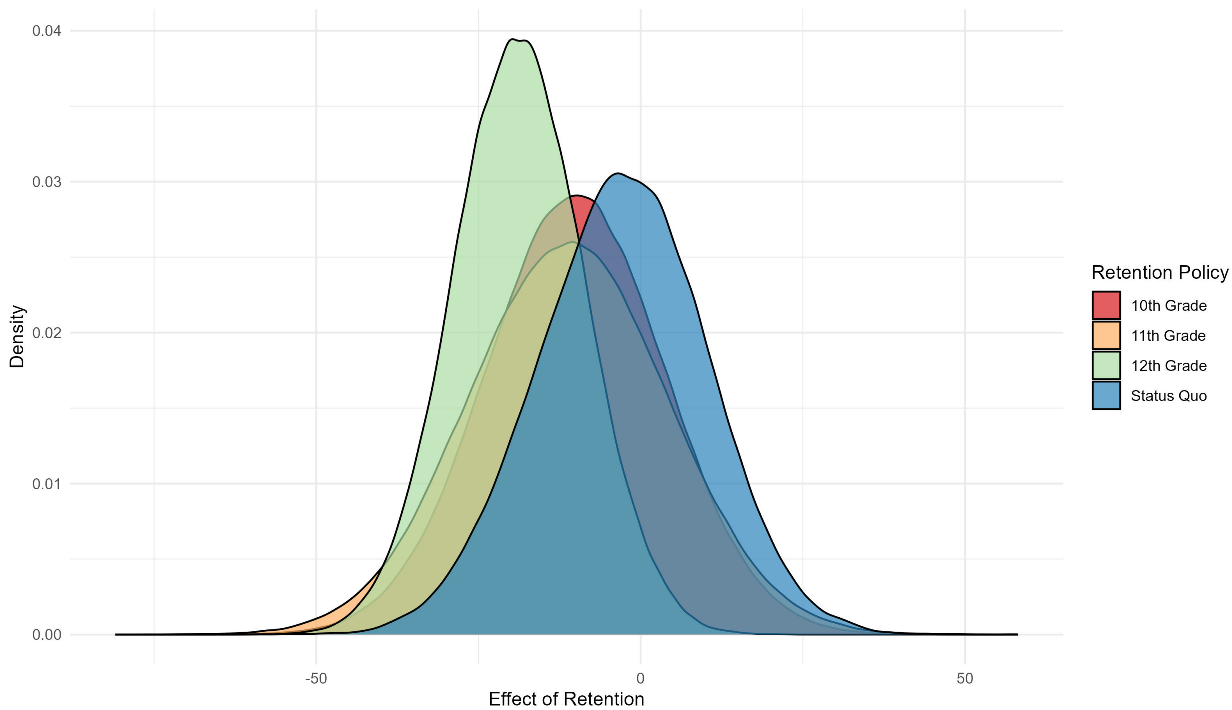
This figure depicts the relationship between the probability of retention and the distance to the cutoff in math in panel (a) and Portuguese in panel (b) The bandwidth for each subject is chosen using the direct plug-in MSE-optimal bandwidth selector on the left hand side of the passing threshold (where retention is variable). This results in a separate bandwidth at each evaluation point. The bandwidth immediately to the left of the threshold is 0.327 for math and 0.355 for Portuguese. Both panels estimate local linear regressions with an Epanechnikov kernel.

Figure C-5: Distributional Impacts on Math Scores, by Subgroup



The figure shows the distribution of math test score impacts for four separate retention policies for the subset of students who graduate in four years. The four policies are the status quo retention policy, and policies that retain all students in grades 10, 11, and 12 respectively. The standard deviation on the 12th grade math exam is 23.6 points.

Figure C-6: Distributional Impacts on Portuguese Scores, by Subgroup



The figure shows the distribution of Portuguese test score impacts for four separate retention policies for the subset of students who graduate in four years. The four policies are the status quo retention policy, and policies that retain all students in grades 10, 11, and 12 respectively. The standard deviation on the 12th grade Portuguese exam is 17.0 points.

D Appendix D: Deriving the Dropout Equation from the Decision Problem of an Optimizing Student

Equation (4) in the main text can be derived from the decision problem of an optimizing student. This derivation is based on a general framework described in Heckman et al. (1999). In deciding whether to drop out, a student may compare her expected future lifetime earnings stream if she stays in school versus if she drops out. Let Y_{1t} and Y_{0t} denote the earnings at time t for a student who drops out versus stays in school. The dropout decision is made sequentially and is based on maximization of expected future earnings net of any psychic costs (C) of attending school:

$$D_{i,t} = 1 \text{ if } \mathbb{E}\left[\sum_{j=0}^{T} \frac{Y_{1,t+j}}{(1+r)^j} - \sum_{j=1}^{T} \frac{Y_{0,t+j}}{(1+r)^j} - C \middle| \Omega_t \right] \geq 0, \text{ else} = 0. \tag{14}$$

The first term is the earnings stream if the person drops out, the second term is the earnings stream if she does not drop out, which reflects the foregone earnings cost of attending an extra year of school, T is the year of retirement, and Ω_t is the information set at time t used to form expectations about future earnings.[27] This framework assumes that individuals forecast their labor market earnings prospects using their own characteristics and using information on the earnings of current labor market participants in the locality where they reside. The dropout decision will also depend on the costs of remaining in school, C, which we assume to be a function of knowledge levels and family background and any effects of prior retentions.

[27] If the individual pursues additional years of schooling, then the earnings could be zero for some years, $Y_{0,t+j} = 0$ for $j = 0, 1, 2, ...$

E Appendix E: Identification of the Covariance Matrix

As described in the text, our model allows for the shocks that affect test scores, retention, and dropout to be jointly correlated, which allows for selection on unobservables. They are assumed to be joint normally distributed:

$$\begin{pmatrix} \varepsilon^M_{i,t}(0) \\ \varepsilon^M_{i,t}(1) \\ \varepsilon^P_{i,t}(0) \\ \varepsilon^P_{i,t}(1) \\ \eta_{i,t+1} \\ \nu_{i,t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma^2_M(0) & - & \sigma_{MP}(0) & - & \sigma_{M\eta}(0) & \sigma_{M\nu}(0) \\ & \sigma^2_M(1) & - & \sigma_{MP}(1) & \sigma_{M\eta}(1) & \sigma_{M\nu}(1) \\ & & \sigma^2_P(0) & - & \sigma_{P\eta}(0) & \sigma_{P\nu}(0) \\ & & & \sigma^2_P(1) & \sigma_{P\eta}(1) & \sigma_{P\nu}(1) \\ & & & & 1 & \sigma_{\eta\nu} \\ & & & & & 1 \end{pmatrix} \right) \tag{15}$$

Shocks that are not contemporaneous are assumed to be independent.

The model is estimated by method-of-moments and all of the moments collectively identify the model parameters. However, it is possible to demonstrate identification of the covariances in (15) by focusing on a subset of histories and moments, as we do here. It is possible to identify $\sigma_{\eta\nu}$, the covariance between shocks in the retention and dropout threshold-crossing models from the moments involving the 10-11-12-12 and 10-11-12-12d histories. These include histories 3, 4, 7, 14, and 15 for STEM students in Table B-2 and histories 3, 4, and 5 for non-STEM students in Table B-3. For ease of interpretability, we focus on the 10-11-12-12 path for non-STEM students. This path corresponds to the following sequence of retention and dropout decisions over four years: $R_1 = 0, D_2 = 0, R_2 = 0, D_3 = 0, R_3 = 1, D_4 = 0, R_4 = 0$.[28] The independence assumptions on the non-contemporaneous errors allow us to decompose the event

$$\{R_1 = 0, D_2 = 0, R_2 = 0, D_3 = 0, R_3 = 1, D_4 = 0, R_0 = 1\} \mid \{Z^R_t\}^4_{t=1}, \{Z^D_t\}^4_{t=2}, \{I_t\}^4_{t=1}, K^M_0, K^P_0, K^P_3, K^P_4$$

[28] Note that, because of the model's notation, D_2 is the first observed dropout decision. Our analysis conditions on students who enroll in secondary school, and D_1 would correspond to the decision to drop out before enrolling.

into the product of three conditionally independent events:

$$\{R_1 = 0, D_2 = 0, R_2 = 0, D_3 = 0\} \mid \{Z_t^R\}_{t=1}^2, \{Z_t^D\}_{t=2}^3, \{I_t\}_{t=1}^2, K_0^M, K_0^P \times$$
$$\{R_3 = 1, D_4 = 0\} \mid Z_3^R, Z_4^D, I_3, K_3^P \times$$
$$\{R_4 = 0\} \mid Z_4^R, K_4^P \ ,$$

where K_0^M and K_0^P are ninth-grade (pre-determined) scores in math and Portuguese, K_3^P is the first observed 12th grade score for the student, and K_4^P is the second observed 12th grade score for the student.

The cross-partial derivative of the probability of the second event with respect to the two exclusion restrictions identifies $\sigma_{\eta\nu}$.

$$\begin{aligned} P(\{R_3 = 1, D_4 = 0\} \mid Z_3^R, Z_4^D, I_3, K_3^P) = P(\nu_3 > \underbrace{-\lambda_0 - \lambda_2 K_3^P - \lambda_3 Z_3^R - \lambda_4 I_3}_{U_3^R}, \\ \eta_4 > \underbrace{-\delta_0 - \delta_2 K_3^P - \delta_3 Z_4^D - \delta_4 - \delta_5 I_3}_{U_4^D}) \\ = \int_{U_3^R}^{\infty} \int_{-\infty}^{U_4^D} f_{\sigma_{\eta\nu}}(\nu, \eta) d\nu d\eta \ , \end{aligned}$$

and

$$\frac{\partial^2 P(\{R_3 = 1, D_4 = 0\} \mid Z_3^R, Z_4^D, I_3, K_3^P)}{\partial Z_3^R \partial Z_4^D} = -f_{\sigma_{\eta\nu}}(U_3^R, U_4^D)\delta_3\lambda_3 \ . \tag{16}$$

With the joint density, $f(\nu, \eta)$, identified, it is then possible to identify any moments from the joint distribution, including $\sigma_{\eta\nu} = \int\int \nu\eta f(\nu, \eta) d\nu d\eta$. When $\sigma_{\eta\nu} = 0$, the partial derivative in (16) equals $-\phi(U_3^R)\phi(U_4^D)\delta_3\lambda_3$. The partial derivative is therefore informative about the value of $\sigma_{\eta\nu}$.[29]

With $\sigma_{\eta\nu}$ identified, we now turn to identification of the remaining parameters of (15) using a control function approach. To economize on notation, we combine the exclusion

[29] As described in Appendix A, our SMM estimation procedure includes moments that are simulated versions of the first order conditions of the likelihood contributions for the histories that we target, where the probability density of a particular secondary school history is simulated rather than computed directly.

restriction and the other covariates in the retention equation into $\mathbf{Z_t^R} := (1, K_t^M, K_t^P, Z_t^R, I_t)$ and similarly the dropout exclusion restriction and covariates in the dropout equation are $\mathbf{Z_t^D} = (1, K_{t-1}^M, K_{t-1}^P, Z_t^D, \sum_{k>0} R_{t-k}, I_{t-1})$. The coefficients in the retention (dropout) equations are combined into the vector λ (δ). The conditional expectation of 12th grade test scores for individuals following the path 10-11-12-12 identifies $\sigma_{\varepsilon\nu}^S(1)$ for S = math, Portuguese. This conditional expectation is

$$\begin{aligned}
&\mathbb{E}[K_4^S(1) \mid R_1 = D_2 = R_2 = D_3 = 0, R_3 = 1, D_4 = 0, R_4 = 0, I_4, K_3^S, \mathbf{Z_4^R}] \\
&\qquad = \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + \mathbb{E}[\varepsilon_4^S(1) \mid R_1 = D_2 = R_2 = D_3 = 0, R_3 = 1, D_4 = 0, R_4 = 0, \mathbf{Z_4^R}] \\
&\qquad = \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + \mathbb{E}[\varepsilon_4^S(1) \mid R_4 = 0, \mathbf{Z_4^R}] \\
&\qquad = \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + \sigma_{\varepsilon\nu}^S(1)\mathbb{E}[\nu_4 \mid \nu_4 \leq -\mathbf{Z_4^R}'\lambda] \\
&\qquad = \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + \sigma_{\varepsilon\nu}^S(1) f(\mathbf{Z_4^R}) \ ,
\end{aligned}$$

where the second equality follows from the independence assumptions. $\sigma_{\varepsilon\nu}^S(1)$ is then identified from the coefficient on $f(\mathbf{Z_4^R})$, which is a parametric control function (the mill's ratio) under our normality assumptions.[30] It follows that the variance of test score shocks in subject S, $\sigma_S^2(1)$, is identified from the squared residuals of this conditional expectation.

The conditional expectation of fourth-year test scores for individuals following the 10-11-12-12-12 and 10-11-12-12-12d paths further identifies $\sigma_{\varepsilon\eta}^S(1)$. This conditional expectation given the path 10-11-12-12-12 is:

$$\begin{aligned}
&\mathbb{E}[K_{i,4}^S(1) \mid R_1 = D_2 = R_2 = D_3 = 0, R_3 = 1, D_4 = 0, R_4 = 1, D_5 = 0, I_4, K_3^S, \mathbf{Z_4^R}, \mathbf{Z_5^D}] \\
&= \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + \mathbb{E}[\varepsilon_4^S(1) \mid R_1 = D_2 = R_2 = D_3 = 0, R_3 = 1, D_4 = 0, R_4 = 1, D_5 = 0, I_4, K_3^S, \mathbf{Z_4^R}, \mathbf{Z_5^D}] \\
&= \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + \mathbb{E}[\varepsilon_4^S(1) \mid R_4 = 1, D_5 = 0, I_4, K_3^S, \mathbf{Z_4^R}, \mathbf{Z_5^D}] \\
&= \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + (\sigma_{\varepsilon\nu}^S(1), \sigma_{\varepsilon\eta}^S(1))\Sigma_{\nu,\eta}^{-1}\mathbb{E}[\nu_4, \eta_5 \mid \nu_4 > -\mathbf{Z_4^R}'\lambda, \eta_5 \leq -\mathbf{Z_5^D}'\delta] \\
&= \gamma_1^S K_3^S + I_4^{S\prime}\beta_1^S + (\sigma_{\varepsilon\nu}^S(1), \sigma_{\varepsilon\eta}^S(1))\Sigma_{\nu,\eta}^{-1}\mathbf{f}(\mathbf{Z_4^R}, \mathbf{Z_5^D}) \ ,
\end{aligned}$$

where the third equality follows from the conditional expectation equaling a linear projection for Normally distributed random variables. $\Sigma_{\nu,\eta}$ is the bottom-right 2X2 submatrix of (15).

[30] The control function could be nonparametrically identified with a continuous exclusion restriction.

Its diagonal elements are normalized to 1 and the off-diagonal element, $\sigma_{\eta\nu}$ has already been identified. Therefore, $\sigma^S_{\varepsilon\nu}(1)$, which has already been identified, and $\sigma^S_{\varepsilon\eta}(1)$ are identified as the coefficients on a vector-valued function of the exclusion restrictions for retention and dropout. Given the assumption of normally distributed unobservables, $\mathbf{f}(\mathbf{Z^R_4}, \mathbf{Z^D_5})$ can be easily simulated using the Geweke– Hajivassiliou–Keane simulator (Geweke, 1989; Hajivassiliou and McFadden, 1998; Keane, 1994).

The final step is to identify $\sigma^S_{\varepsilon\nu}(0)$, $\sigma^S_{\varepsilon\eta}(0)$, and $\sigma^2_S(0)$. These can be identified from the coefficients on a control function estimator for the conditional expectation of K^S_3 for people who are not retained and do not drop out. Consider individuals who follow path 10-11-12 in subject S. Then

$$\begin{aligned}
&\mathbb{E}[K^S_3 \mid R_1 = D_2 = R_2 = D_3 = R_3 = 0, \{I_t\}^3_{t=1}, \{\mathbf{Z^R_t}\}^3_{t=1}, \{\mathbf{Z^D_t}\}^3_{t=2}] \\
&\quad = {\gamma_0}^3 K_{i,t} + ({\gamma_0}^2 I_{i,t+1} + \gamma_0 I_{i,t+2} + I_{i,t+3})'\beta_0 + \mathbb{E}[\varepsilon_{i,t+3}(0) \mid R_3 = 0, \mathbf{Z^R_3}] \\
&\quad + \gamma_0 \mathbb{E}[\varepsilon_{i,t+2}(0) \mid R_2 = D_3 = 0, \mathbf{Z^R_2}, \mathbf{Z^D_3}] \\
&\quad + {\gamma_0}^2 \mathbb{E}[\varepsilon_{i,t+1}(0) \mid R_1 = D_2 = 0, \mathbf{Z^R_1}, \mathbf{Z^D_2}] \ ,
\end{aligned} \tag{17}$$

where the independence assumptions on the errors have been imposed in the equality. We have the following results for the three control functions:

$$\begin{aligned}
\mathbb{E}[\varepsilon_{i,t+3}(0) \mid R_3 = 0, \mathbf{Z^R_3}] &= \sigma^S_{\varepsilon\nu}(0)\mathbb{E}[\nu_3 \mid \nu_3 \leq -\mathbf{Z^R_3}'\lambda] \\
\mathbb{E}[\varepsilon_{i,t+2}(0) \mid R_2 = D_3 = 0, \mathbf{Z^R_2}, \mathbf{Z^D_3}] &= (\sigma^S_{\varepsilon\nu}(0), \sigma^S_{\varepsilon\eta}(0))\Sigma^{-1}_{\nu,\eta}\mathbb{E}[\nu_2, \eta_3 \mid \nu_2 \leq -\mathbf{Z^R_2}'\lambda, \eta_3 \leq -\mathbf{Z^D_3}'\delta] \\
\mathbb{E}[\varepsilon_{i,t+1}(0) \mid R_1 = D_2 = 0, \mathbf{Z^R_1}, \mathbf{Z^D_2}] &= (\sigma^S_{\varepsilon\nu}(0), \sigma^S_{\varepsilon\eta}(0))\Sigma^{-1}_{\nu,\eta}\mathbb{E}[\nu_1, \eta_2 \mid \nu_1 \leq -\mathbf{Z^R_1}'\lambda, \eta_2 \leq -\mathbf{Z^D_2}'\delta] \ .
\end{aligned}$$

This means that the control function estimator for subject S in (17) contains over-identifying information to identify $\sigma^S_{\varepsilon\nu}$ and $\sigma^S_{\varepsilon\eta}(0)$, as there are three separate two-dimensional control functions that each vary with the time-varying exclusion restrictions, and there are only two coefficients to identify. The variance of test score shocks for each subject is then identified by the squared residuals, $K^S_3 - \mathbb{E}[K^S_3 \mid R_1 = D_2 = R_2 = D_3 = R_3 = 0, \{I_t\}^3_{t=1}, \{\mathbf{Z^R_t}\}^3_{t=1}, \{\mathbf{Z^D_t}\}^3_{t=2}]$.