IDB WORKING PAPER SERIES Nº IDB-WP-758

Retaking in High Stakes Exams:

Is Less More?

Kala Krishna
Sergey Lychagin
Verónica Frisancho

Inter-American Development Bank
Department of Research and Chief Economist

April 2017

Retaking in High Stakes Exams:

Is Less More?

Kala Krishna*
Sergey Lychagin**
Verónica Frisancho***

* Pennsylvania State University, CES-Ifo and National Bureau of Economic Research
** Central European University
*** Inter-American Development Bank

April 2017

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Krishna, Kala.
Retaking in high stakes exams: is less more? / Kala Krishna, Sergey Lychagin, Verónica Frisancho.
p. cm. — (IDB Working Paper Series ; 758)
Includes bibliographic references.
1. College entrance achievement tests-Turkey-Econometric models. 2. Examinations-Psychological aspects-Econometric models. I. Lychagin, Sergey. II. Frisancho Robles, Verónica C. III. Inter-American Development Bank. Department of Research and Chief Economist. IV. Title. V. Series.
IDB-WP-758

http://www.iadb.org

Abstract

Placement, both in university and in the civil service, according to performance in competitive exams is the norm in much of the world. Repeat taking of such exams is common despite the private and social costs it imposes. We develop and estimate a structural model of exam retaking using data from Turkey's university placement exam. Limiting retaking results in all agents gaining ex-ante, and most gaining ex-post. This result comes from a general equilibrium effect: retakers crowd the market and impose negative spillovers on others by raising acceptance cutoffs.

Keywords: Repeat taking, Higher education, Dynamic choice, General equilibrium

JEL Classification: C38, C61, C63, I24, I26

1 Introduction

In much of the world, both now and in the past, competitive exams have been used to select the best and brightest. The examination required to be chosen as a civil servant in Imperial China is a classic example. Such exams remain common in many countries, including China, Japan, India, the United Kingdom and the United States. Admission to university in many countries is also similarly structured and is fiercely competitive. Students often retake these exams many times, spending enormous amounts of time, money and effort in the hopes of a better placement. In Korea, for example, almost 18 billion dollars were spent in 2013 in prep schools by those taking the college entry examination. Students spent so much time studying (15 hour days are the norm) that the government had to order prep schools to close by 10 in the evening. There have even been suicides among students who perform poorly. Despite such extreme duress, 20 percent retake the exam in hopes of doing better.[1] In China, the infamous "gaokao" taken to enter university creates extreme stress.[2]

Retaking has both positive and negative elements: on the plus side, retaking reduces the impact of bad luck as it insures against downside risk. In addition, as those who underperform will retake, the extent of student-college mismatch could be reduced as well. It may also help level the playing field if disadvantaged students learn more upon retaking. On the minus side, retaking tends to be excessive, as the individual's gains from moving up in the rankings necessarily come at the cost of others.[3] Retakers increase competition for a given number of slots, which has general equilibrium consequences: admission standards rise with more students competing for seats. This effect is magnified when the quality of colleges is skewed and the best ones are subsidized. In this case, there are considerable rents to getting into elite schools, but these are dissipated through excessive effort and retaking.

In the United States, standardized exams are studied for quite intensely, and taken multiple times, especially by better-off students. Retaking is also related to the issue of "red-shirting."[4] In that country children, especially boys, often start school a year late in the hope that this will allow them to do better then their peers. Deming and Dynarski (2008) provide a lucid summary of work on this topic in several settings.

Given the prevalence of retaking it is surprising that, at least to our knowledge, there is no systematic analysis of its costs and benefits in a general equilibrium setting. The only previous

[1] See the article entitled "Trading Delayed as 650,000 South Koreans Take College Test," Bloomberg News, November 6, 2013. Also see the article "South Korea's dreaded college entrance exam is the stuff of high school nightmares, but is it producing "robots"?", CBS News, November 7, 2013.

[2] http://www.huffingtonpost.com/2012/07/02/china-test_n_1644306.html

[3] Due to this externality, retaking is likely to be excessive as in Akerlof (1976). Though much of the contest literature views effort as desirable, the negative spillovers suggest that when effort is exerted only to improve standing it is socially costly.

[4] This refers to holding an athlete back until he is stronger and more able to compete.

work on the topic, Vigdor and Clotfelter (2003) restricts attention to partial equilibrium which assumes away the key externality at work. This paper addresses this deficit and builds and estimates a structural dynamic model of retaking where forward looking students choose whether or not to retake the exam by weighing the expected benefits in terms of their future score with the costs of retaking. The model lets us answer the following questions that are key for policy: if retaking were limited, or even eliminated, and students had to take the exam at the end of high school, what would be the consequences? Who would gain and who would lose? Is it possible to change the system so that most people gain from the change in steady state?

We rely on 2002 data on the Turkish college admission exam. Only about a third of the exam takers in Turkey were taking the placement exam for the first time, while roughly 10 percent of them were on at least their fourth attempt. Though there are roughly as many seats as there are high school graduates, the large share of retakers creates an overhang.[5] The Turkish case is the ideal setting for our purposes due to several features: high-stakes admission exams; relatively clear rules; and stability of the system in the years prior to 2002, both in terms of the number of high school graduates and exam takers and the number of seats available.

We are able to estimate the structural parameters of the dynamic model: retaking costs, utility of placement, and learning between attempts. We allow all parameters to vary across income groups since their costs and benefits from retaking are likely to be different. As we only have cross-sectional data on first-time and repeat takers, we cannot rely on standard dynamic estimation methods. In particular, identifying selection into retaking and improvement in scores between attempts is especially difficult in our case. To separate selection from learning we use the fact that high school grade point average (GPA) is predetermined and thus unaffected by learning. Therefore, the distribution of high school GPA[6] of retakers captures selection. On the other hand, exam performance of retakers reflects both selection and learning. By looking at the joint distribution of the two performance measures across attempts we can tease out learning from selection. However, this requires additional assumptions. In particular, we need to assume stationarity of the equilibrium as well as stationarity of the population and its composition. We find that, on average, low-ability students have a higher probability of retaking, which moves the score distribution of retakers to the left relative to that of first-time takers. As a result, not controlling for selection underestimates learning.

We find that more advantaged students tend to have lower costs of retaking. Utility of placement differs mostly for the best schools, which the poor seem to value far more than the

[5] It is worth noting that increasing the number of seats, as seems to be suggested in Hatakenaka (2006), may not solve the problem, as retaking is an equilibrium phenomenon.

[6] High school GPA is notoriously hard to compare across schools due to differential grading standards. However, as we also have data on performance in a common exam, the raw GPA can be rescaled to be made comparable. We discuss the two different ways of rescaling GPA in Appendix A.1.

rich. This makes sense, as the rich have access to costly outside options in the international market that the poor cannot afford. Learning gains are between 0.2 and 0.5 standard deviations of the placement score and are highest in the middle income group and lowest among the rich. This is consistent with the rich having already reached their performance frontier by the first attempt.

An advantage of modeling equilibrium and estimating the structural parameters of the model as done here is that we can perform counterfactual experiments with the aim of guiding policy. In steady state, most students tend to gain from banning retaking. Though each student is worse off by not being allowed to retake for given cutoff scores, banning retaking makes cutoff scores fall as competition for placement is less fierce. This occurs both because fewer students compete for placement at any time and because there is none of the learning that can occur with retaking. For our estimated model, the counterfactual simulations show that this general equilibrium effect dominates, so that everyone is ex ante better off by restricting retaking. Even ex post, about 80 percent gain. Nevertheless, if students are naive and cannot anticipate general equilibrium effects of such reforms, they will resist the restrictions. It is worth emphasizing that the welfare effects of discouraging retaking depend on the estimated parameters. Appendix A.4 provides two examples that look quite similar in terms of the parameters, but that differ drastically in terms of the effects of higher retaking costs.

While our model captures the essence of the issues we choose to focus on, several limitations need to be pointed out. First, retaking is expected to improve matches between students and schools by giving second chances. In this paper, we do not postulate any gains, private or social, from assortative matching since we cannot identify them in our data. We are, however, able to capture the extent to which students are under-placed under various counterfactual scenarios.

Second, our base model does not account for endogenous effort. However, in an extension (in Appendix A.3) we are able to account for endogenous pre-exam investment in a limited way. We model costs incurred in high school by allowing students to choose among three high school types and to pay for extra tutoring. Our data do not allow us, however, to say much about effort between exam attempts. We estimate these fixed costs and allow them to vary by income group and number of attempts and interpret them as including any effort costs as well as psychic costs or time forgone.[7]

Third, though we find that students learn between retaking attempts, we do not include any benefits of learning (such as higher wages later on in life) other than those that operate via placement. We do this because we have no data on the extent of such benefits. If the absolute level of utility depended on the score, there could be further benefits from retaking not captured in our model.

[7] This is not a problem as far as the estimates go, but is a potential problem for conducting counterfactuals as these retaking cost estimates that capture effort expended between attempts are subject to the Lucas Critique.

Fourth, in our estimation we assume that preferences are purely vertical, though they can differ by income. The high retaking rates observed are also consistent with strongly vertical preferences. While this assumption captures what seems to be a clear hierarchy among schools, it assumes away factors such as idiosyncratic preferences across majors or location.

As mentioned above, there are only a handful of papers that look at the issue of retaking. Vigdor and Clotfelter (2003) look at retaking the SATs in the United States. They calibrate a partial equilibrium model and show that the practice of using the best SAT score serves to discriminate in favor of more advantaged groups, as these have lower costs of retaking, retake more often, and so get higher maximum scores across attempts. However, as they do not model the equilibrium, they are forced to assume that schools do not change their admission rules in their counterfactuals. We show that these general equilibrium effects are critical; had we made the same assumption as them, we would have mistakenly found that banning retaking was unambiguously bad.

Törnkvist and Henriksson (2004) also look at retaking and learning. They use data on SweSAT, the Swedish version of the SAT, and document patterns in it. As in the United States, the SweSAT is one of many criteria that universities use in granting admission. It is offered biannually and taken multiple times, as the best score obtained is used. Using panel data on four consecutive rounds of the exam they follow students and so are able to pin down learning and how it varies across groups. They also find learning gains, especially in the second attempt. They find some evidence of differential learning gains across income groups, but these are not robust. They document that richer and higher- ability students have higher retaking rates. To our knowledge, ours is the first paper that estimates a structural model of retaking.

Methodologically, we build upon the estimator developed by Hotz and Miller (1993). Their approach relies on having data on agent actions and state transitions. We extend this approach to a cross-sectional dataset in which state transitions are not observed. Our work is tangentially related to the literature on contests. However, in contrast to our model, this literature is explicitly strategic and focuses on small numbers interactions. Much of it focuses on how to elicit more effort from agents as effort is what the principal cares about.[8] Our paper models the contest as an anonymous game where effort is not per se desirable and students take cutoffs to be admitted as given. This is analogous to monopolistic competition where firms take the price index as given.[9]

In what follows, we first lay out the data and a simple model that captures the essential aspects of the Turkish system. In Section 4 we discuss the intuition behind the model's identification

[8] For instance, see Fu (2006) and Fain (2009).

[9] In fact, retaking being excessive in our model is analogous to the result of Mankiw and Whinston (1986) on excessive entry with homogeneous firms and monopolistic competition. Just as each firm does not internalize the effect of its entry on the profits of existing firms and this profit stealing effect results in excessive entry, students who retake do not internalize the effect of their retaking on the placement of other students.

and our estimation procedure. We report the estimation results in Section 5. Section 6 contains the counterfactual exercises. Section 7 concludes.

2 Institutional Setting and Data

Turkey has a highly centralized college admission procedure. All potential college applicants in a given year have to take the ÖSS, which is used for college placement and simultaneously administered all over the country once a year by OSYM (Student Selection and Placement Center). In what follows, we outline the rules that applied to college admissions in 2002 and describe our data.

2.1 Turkish College Admission Exam in 2002

The exam is composed of multiple choice questions with negative marking for incorrect answers. Student i's performance is evaluated in four subjects: Mathematics, Turkish, Science, and Social Studies. These subject scores, s_{ij}, together with the normalized high school GPA, g_i, are used to construct placement scores, s_i:

$$s_i = w_{ig} g_i + \sum_{j \in \{M,T,Sc,SS\}} w_j s_{ij}. \tag{1}$$

There are three sets of subject weights w_j, so that each student receives three placement scores.[10] Admission to each college program is based on one of these scores depending on the program's major.[11] The weight on GPA, w_{ig}, is student-specific. Students are encouraged to apply to college programs compatible with their high school tracks and discouraged from retaking after placement.[12]

After taking the placement exam and learning the results, students are provided with information on the selectivity of available programs. Then, they submit their college preferences.[13] Placement is based on a serial dictatorship algorithm: a student is placed in his most preferred program, conditional on the availability of seats after all the applicants with higher scores are placed.

[10] The abbreviated names of these scores are Y-ÖSS-SAY, Y-ÖSS-SÖZ and Y-ÖSS-EA. Students who take an optional language test receive four scores. The fourth score, Y-ÖSS-DİL, is used for selection into language programs.

[11] For instance, programs in engineering accept the score called Y-ÖSS-SAY, which puts higher weight on mathematics and science, while programs in anthropology use Y-ÖSS-EA, which puts more weight on the Turkish and the social studies components.

[12] For example, if a student from the social studies track applies to an engineering program, this program uses $w_{ig} = 0.2$ to calculate the student's score. However, if the student comes from the science track, then $w_{ig} = 0.5$. The weight on GPA is further reduced by 50 percent if the student was placed to a college program in the previous year.

[13] In addition to their scores, students receive a booklet with previous year's cutoff scores for each program (i.e., the score of the last student admitted). Cutoff scores in the most popular programs are very stable across years.

Students are not placed if their scores are below a threshold or if all their chosen programs are unavailable to them (i.e., they are filled up by better-performing students). Such students can retake the exam without penalty. Students who are placed are also allowed to retake; however, their placement scores are heavily penalized if they retake the following year. Students can always avoid unacceptable placements by submitting blank preferences.

2.2 Science Students and the Market They Target

Modeling exam taking for all potential college applicants is hardly feasible given our data. With few exceptions, any person who finished high school in the past is allowed to take the exam. This includes current college students, college graduates who wish to obtain a second degree, high school graduates who chose employment over college in the past, students who apply to programs incompatible with their high school tracks, and so on. If we were to model exam-taking in the entire population, our model would have to capture labor market participation, the choice of high school track and college major, and a decision to stay or drop out from college in a dynamic setting. All this is well beyond the scope of our paper.

To circumvent such issues, we identify a segment of the market for college seats that is fairly isolated from the rest of the system. The demand side of this market consists of students graduating in the science track from Turkish high schools who are not enrolled in a college program. The supply side of the market comes from those programs that use the Y-ÖSS-SAY score in their placements. Thus, we assume that science track students target programs accepting Y-ÖSS-SAY. We treat this group as if it exists in complete isolation from the rest of the population, ruling out any effects that they have on applicants from other tracks and vice versa. We do not drop students on the basis of where they apply.[14]

We choose this as a natural market to study for the following reasons. First, the Turkish system provides incentives for students to build on their high school specialization. Applying to a college major that does not match one's high school track results in the student's GPA weight being reduced from 0.5 to 0.2.[15]

Table 1 demonstrates how segmented the market for seats is. Less than one percent of placements among the science track students account for programs that are incompatible with the science track and have competitive admissions. Virtually no students from the other two major aca-

[14] Our central argument that retaking has a general equilibrium effect on welfare via the admission cutoffs requires that admissions are competitive. This is why we focus on the science track. Science track students are more likely to apply and be placed to programs with competitive admissions than, for example, those who specialize in social studies or Turkish and mathematics, the other two major high school tracks (see Table 1).

[15] Given that the normalized GPA ranges between 30 and 80 points, applying to a program not matching the student's track sets him back by 9-24 points, which is roughly one third to one standard deviation of the placement score in our sample. For this reason, while students occasionally apply to programs incompatible with their tracks, very few such applications are successful.

Percentage of placements, by student's track	High school track		
College programs:	Science	SS	TM
With a non-binding quota	24%	71%	40%
With a binding quota			
compatible w/science track, incompatible w/SS and TM			
accepting Y-ÖSS-SAY	51%	0%	0%
compatible w/science track, compatible with SS or TM			
accepting Y-ÖSS-SAY	11%	0%	0%
accepting scores other than Y-ÖSS-SAY	14%	2%	26%
incompatible with the science track	1%	27%	34%
Total:	100%	100%	100%

Table 1. Placements within Science, Social Studies and Turkish-Math High School Tracks, by Type of College Program

demic tracks (Social Studies and Turkish-Math) are admitted to programs that are only compatible with the science track or that use Y-ÖSS-SAY scores for admission.

There are, however, some notable exceptions from strict tracking by subject. Many programs in economics, business and education are compatible with multiple high school tracks and admit students based on scores other than Y-ÖSS-SAY. According to Table 1, up to 14 percent of science track students are placed to such programs. Yet, despite these exceptions, most science track students are not competing for placement with those outside the science track. In our market definition, we exclude all enrolled students from the demand side. This is justified by the penalty imposed on placed retakers.[16]

While we observe in the data that 14 percent of exam takers are currently attending college,[17] they tend to either target different programs or retake without much hope to succeed. As shown in Table 2, of those who were placed in 2002, 91 percent were not attending college. An overwhelming majority of these (72/91) were placed in selective programs. On the other hand, 9 percent were attending college when placed, but these students were far less likely (5/9) to go to programs with competitive placement.

Dropping college students substantially simplifies our estimation approach: if students are allowed to retake after placement, their current state would include all their past placements, which

[16] If a student retakes the exam in the year after a successful placement into a bachelor program, the weight on GPA, w_{ig}, is automatically halved. After spending two years in college, relatively few students should find it worthwhile to start over in a new program.

[17] We cannot directly calculate the probability of retaking conditional on placement as we do not track students after they are enrolled. The number of retakers in college in our original sample is 6,028 out of 42,731. Assuming that the number of students in each year of college is the same as those placed in the 2002 cohort (15,500 students) and each student spends four years in college we would have $15,500 \times 4 = 62,000$ students at risk of retaking. This gives roughly $6,028/62,000 = 9.7$ percent as a back-of-the-envelope estimate.

Percentage of students placed in 2002	Program of placement: non-binding quota	binding quota	Total
Not attending college	19%	72%	91%
Attending college	4%	5%	9%
Total	23%	77%	100%

Table 2. Placements of Science Track Applicants by College Enrollment Status at the Time of Taking the Exam.

we do not observe in the data.[18] If we ignore retakers who were previously placed, we can perfectly back out every applicant's history of actions: if someone is taking the exam for the tth time, it means that he chose to retake in the previous $t-1$ attempts. Being able to know the history of each student's actions is crucial in our identification strategy, as we have to control for selection on student ability caused by decisions to retake.

Focusing on a specific group of students, of course, imposes some restrictions on external validity of our estimates. First, as we demonstrate later on, the predicted retaking rates and welfare measures depend on the intensity of competition for seats. As Table 1 demonstrates, many more science students pursue competitive programs than students from the other popular tracks. Thus, our results may not be directly applicable to other segments of the Turkish student population. Second, although high school graduates cannot easily apply to a program compatible with a different track, any policy change affecting college admissions may induce some students to choose different tracks while in high school. In our counterfactuals, we do not account for this adjustment channel as we do not have the detailed data needed to set up and estimate a model of high school track choice.

Our modeling and estimation approach relies on the assumption that the system is in steady state.

Assumption 1. ***(Stationarity)*** *The population of exam takers and the supply of seats do not change over time and admission cutoffs are stable.*

Table 3 documents the trends in demand and supply for the system as a whole. The number of high school graduates and the number of exam takers was roughly stable between 2000 and 2003. There was a slight increase in admissions to four-year programs. These aggregate numbers are for a larger market than what we focus on. Unfortunately, more detailed data for our market segment were not available.

[18] For instance, suppose one observes an enrolled student in his third exam attempt. Given our data, it is impossible to find the sequence of past actions that led to this outcome. The student could have been placed after his first attempt and then taken the exam twice while in college. Or he could have been placed only after his second attempt.

Year	High school graduates	ÖSS exam takers	Admissions to 4+ year programs
2000	536,124	1,414,823	164,977
2001	532,952	1,473,908	170,473
2002	507,363	1,540,422	176,612
2003	530,259	1,502,605	187,345

Notes: The number of high school graduates includes all individuals who graduated from formal secondary education institutions in a given year; data source — Education Statistics of Turkey 2005–2006. The number of exam takers includes those who applied to take the college admissions exam; data source — Türk Yükseköğretiminin Bugünkü Durumu 2003, 2005. Admissions: 4+ year programs include regular and evening university programs, but exclude distance education and associate degrees. Data source: Higher Education Statistics, years 2000–2004.

Table 3. Trends in College Applications and Admissions

2.3 The Data

Our data cover a random sample of about 42,731 students who took the ÖSS in 2002 and who were in the science track. ÖSYM data comes from three sources: students' application forms, a survey given in 2002, and administrative data on high school GPA and scores in each part of the exam. After cleaning the data, dealing with some minor inconsistencies (4 percent) across different data sources, and dropping those who retake while already enrolled in a university program (13 percent), as well as those with missing data (8 percent) we lose roughly 25 percent of the observations. We restrict our attention to the 31,554 from the science track that remain.

For each student, our database contains information on high school characteristics (type of school), high school GPA, standing at the time of the exam (high school senior, repeat taker), individual and background characteristics (gender, household income, parents' education and occupation, family size, time and money spent on private tutoring, and number of previous attempts), and performance on the exam (both the raw and weighted scores).[19]

It is worth noting that the GPA in our dataset does not account for quality differences across schools or grade inflation. While the exam authority corrects for this by using the normalized GPA, g_i, the latter is neither reported in the data, nor can it be easily computed from the available inputs. We use an approximation to g_i that relies on available information as described in Appendix A.1, where we also reconstruct the official g_i for a subset of students as a consistency check.

[19] We could find only a few papers that have explored the Turkish data set. Tansel and Bircan (2005) studies the determinants of attendance at private tutoring centers and its effects on performance. Saygin (2016) looks at the gender gap in college. Moreover, Caner and Okten (2010) looks at career choice using data on preferences, while Caner and Okten (2013) examines how the benefits of publicly subsidized higher education are distributed among students with different socioeconomic backgrounds.

2.4 Preliminary Evidence on Retaking

Despite the fact that retaking requires a year of waiting and preparation, this phenomenon is highly prevalent in Turkey. In 2002, more than 50 percent of the science track applicants were repeat takers.[20] According to our data, approximately 80 percent of retakers are not employed at the time of the exam.

High retaking rates could arise from four sources: a low cost of retaking, a high value of a better placement, a probable improvement in scores due to learning, and uncertainty in test results. While the first three mechanisms clearly increase retaking, the fourth one is more subtle. If exam outcomes are highly uncertain, retaking is akin to gambling with a limit on losses: exam takers are fully exposed to lucky draws of placement scores while being insulated from the bad ones by the option to retake. Insured against bad shocks, students retake more if they are exposed to higher uncertainty in future scores.[21]

How prevalent is retaking in different socioeconomic groups? Frisancho, Krishna, Lychagin, and Yavas (2016) suggest that the disadvantaged have greater learning gains than the advantaged. This suggests that studying for the exam has decreasing returns; students from well-off families get more private tutoring by the time of the first attempt than low-income students do. Consequently, we would expect low-income students to retake more often. On the other hand, if disadvantaged students have higher costs of retaking, then they will be less likely to retake.

The incentives to retake also depend on the gains from a higher score: if the value of doing better rises sharply at a particular score level, then students with scores just below this level are more likely to retake. For example, if the best school is far superior to the second best, students just below the best school's admission cutoff score should tend to retake more often. This mechanism makes retaking rates depend on student ability and the structure of placement payoffs.

Table 4 shows the number of students in low, middle and high income groups for each retaking attempt.[22] The data suggests that the poor are increasingly prevalent among serial retakers. This is consistent with higher learning gains, lower costs of retaking, or, to the extent that the poor also tend to be more present at lower scores, that the value of a better school increases considerably at lower scores. In either case, this preliminary evidence suggests that eliminating or restricting retaking may affect the poor more adversely than other groups.

---

[20] These numbers are much higher in the social studies track. Overall, about 67 percent are retakers.

[21] This is conceptually similar to a well-known result from the search theory that an increase in the variance of prices increases time and effort invested in bargain hunting (e.g., see Stigler (1961)).

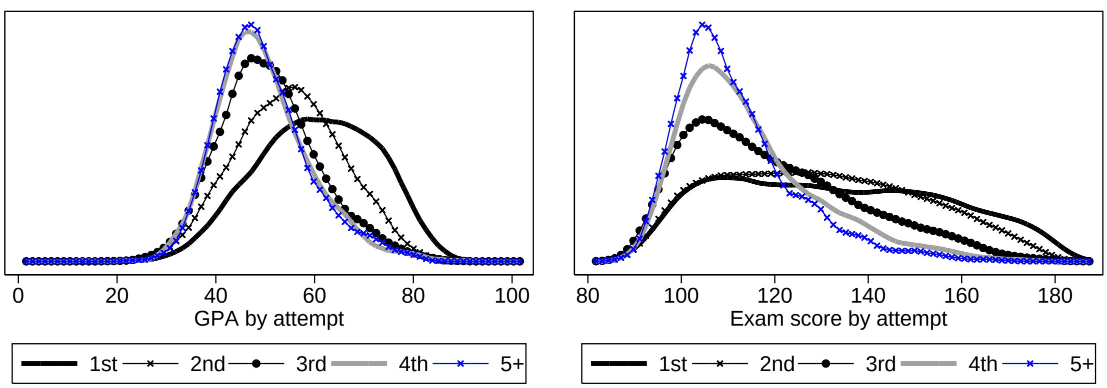
[22] Our definition of income groups splits the population into three roughly equal parts. Students in the low income group report monthly household income of less than 250 Turkish lira (YTL). Households earning more than 500 YTL are classified as high-income ones. Those in between 250 YTL and 500 YTL are middle-income households. The socioeconomic data is relatively coarse (interval data are reported) and there is an incentive to under-report incomes as scholarships are income-based. We expect the order to be more correct than the level reported and this is why we use this coarse grouping.

Income	Low	Medium	High	Total
Attempt				
1	4,454	6,388	4,757	15,599
	29%	41%	30%	100%
2	2,635	3,221	2,043	7,899
	33%	41%	26%	100%
3	1,547	1,681	917	4,145
	37%	41%	22%	100%
4	861	929	401	2,191
	39%	42%	18%	100%
5+	699	666	355	1,720
	41%	39%	21%	100%

Notes: The first row of each cell contains the total number of exam takers in a given attempt and income group. The second row shows the percentage of each income group by attempt.

Table 4. Number and Percentage of Exam Takers by Attempt and Income

Figure 1. Distributions of Exam Scores and High School GPA by Attempt



(a) High school GPA

(b) Exam scores

Figure 1a shows the distribution of high school GPA across the number of attempts. As is evident, the distribution moves to the left, suggesting that weaker students face greater gains/lower costs of retaking and thus tend to retake more often. Figure 1b plots the empirical distributions of exam scores by number of attempts. The distribution of scores shifts to the left as well, consistent with negative selection into retaking (movement of the distribution to the left) dominating learning (movement of the distribution to the right).

3 Modeling the Turkish System

In this section, we describe the components of our model. First, we set up the factor model that governs the high school GPA and exam score realizations in the population of first-time takers. Following this, we lay out a theory explaining demand for seats and retaking decisions in a steady state equilibrium.[23]

3.1 Performance in High School and on the Entrance Exam

There is a continuum of students. We postulate that the normalized high school GPA[24] for student i is given by

$$g_i = X_i'\beta_g(I_i) + \theta_i'\alpha_g(I_i) + \varepsilon_{ig}. \tag{2}$$

where X_i is a vector of individual characteristics (laid out in Table 5) that do not vary in time and are potentially correlated with the student's ability. The remaining terms, $\theta_i'\alpha_g(I_i) + \varepsilon_{ig}$, constitute the residual; $\theta_i = [\theta_{iq}, \theta_{iv}]'$ represents the unobserved part of quantitative and verbal ability, which affects the student's performance in all settings. The components of this ability vector, θ_{iq} and θ_{iv}, are allowed to be correlated. If more able students are likely to do better in both verbal and quantitative tasks, this correlation will be positive. θ_i is observed by the student, but unobserved by the econometrician. The shock, ε_{ig}, captures the randomness associated with the GPA.

The subject scores on the t^{th} attempt are

$$s_{ijt} = X_i'\beta_j(I_i) + \theta_i'\alpha_j(I_i) + \sum_{\tau=2}^{t} \lambda_{ij\tau} + \varepsilon_{ijt}, \quad j \in \{M, T, Sc, SS\} \tag{3}$$

where ε_{ijt} is the shock to the student's score in subject j (math, Turkish, science and social studies) and attempt t, and $\lambda_{ij\tau}$ captures improvement in student's performance between

---

[23] In Appendix A.3, we set up an extension of this model, which allows for pre-test investments prior to the first attempt. We do so because banning retaking could intensify the rat race in high school.

[24] To account for differential grading practices across schools, GPAs are normalized to be comparable using the school's performance in the university entrance exam. Our normalization differs slightly from the official one due to data limitations; see Appendix A.1 for details.

Income	Low	Medium	High
Father's education			
Primary or less	6,567	4,305	1,320
Middle/High school	2,538	4,702	2,392
2-year higher education	136	1,014	812
College/Master/Phd	245	1,958	3,463
Missing	710	906	486
Mother's education			
Primary or less	8,658	8,656	2,702
Middle/High school	1,009	2,921	2,756
2-year higher education	26	307	973
College/Master/Phd	32	285	1,615
Missing	471	716	427
Father's occupation			
Employer	42	299	847
Works for wages/salary	3,847	8,474	5,569
Self-employed	3,506	3,085	1,705
Unemployed/not in LF	2,801	1,027	352
Mother's occupation			
Employer	4	23	104
Works for wages/salary	304	1,456	3,249
Self-employed	534	355	331
Unemployed/not in LF	9,354	11,051	4,789
Expenditures on prep. schools			
No prep school	2,350	1,603	423
Scholarship	410	395	245
Less than 1b	3,790	5,211	2,137
1-2 b	906	2,936	2,790
More than 2b	219	776	2,203
Missing	2,521	1,964	675

Income	Low	Medium	High
Funds for college			
Family/rental inc/scholarship	3,044	5,164	4,096
Work	2,633	2,937	1,894
Loan	3,837	4,127	2,002
Other	682	657	481
High school city population			
<10k	1,345	971	263
10-50k	1,754	1,793	766
50-250k	2,487	2,913	1,602
250-1000k	1,637	2,284	1,374
>1million	2,289	4,192	4,108
Missing	684	732	360
Family size			
1-2 children	4,504	8,446	6,999
>3 children	5,692	4,439	1,474
Gender			
Female	4,185	5,761	3,921
Male	6,011	7,124	4,552
High school category			
Public	6,951	7,262	3,326
Private	1,447	2,480	2,028
Anatolian/Science	1,301	2,696	2,973
Other	497	447	146
Access to internet			
Yes, at home	296	1,316	2,698
Yes, not at home	2,733	4,455	2,911
No	6,792	6,681	2,666
Missing	375	433	198

Notes: Each cell reports the number of students who share the respective characteristic. Family income: low (< 250 YTL monthly; approx. USD 375), medium (250–500 YTL), high (> 500 YTL).

Table 5. Demographic Composition of Exam Takers from the Science Track

attempts $\tau - 1$ and τ. The effect of $\lambda_{ij\tau}$ on i's score accumulates over attempts, while that of ε_{ijt} is transitory.

Individual characteristics in equations (2) and (3), X_i, capture variation in student ability related to observable socioeconomic factors and earlier schooling choices. For instance, exam takers from larger cities tend to score higher. While we cannot identify the exact mechanism behind this correlation, it may reflect regional differences in access to good schools. Pre-exam investment is likely to play an important role, too: exam takers who attended preparatory courses tend to perform far better than those who did not. We also include parental education and occupational status as the part of X_i, since these variables may correlate with early childhood investments not captured by the school controls.

The raw data suggest that the correlation between performance and the elements of X_i varies substantially with income. For instance, males tend to score higher than females, and this gender gap is decreasing in income.[25] To allow for such asymmetries, we split the population by income into three groups of roughly equal size and introduce an income group indicator I_i that takes three values (low, medium and high income). We allow all parameters of the model, including α and β from equations (2) and (3), to depend on I_i without any restrictions.

Student i's placement score in attempt t is defined by equation (1): it is a weighted sum of subject scores in that same attempt and the high school GPA. In what follows, it is convenient to split the score into two components, $\bar{s}_{it}$ and ε_{it}:

$$\begin{aligned} s_{it} &= w_{ig} g_i + \sum_j w_j s_{ijt} \\ &= w_{ig} g_i + X_i' \sum_j w_j \beta_j(I_i) + \theta_i' \sum_j w_j \alpha_j(I_i) + \sum_j \sum_{\tau=2}^{t} w_j \lambda_{ij\tau} + \sum_j w_j \varepsilon_{ijt} \\ &= w_{ig} g_i + X_i' \beta(I_i) + \theta_i' \alpha(I_i) + \sum_{\tau=2}^{t} \lambda_{i\tau} + \varepsilon_{it} = \bar{s}_{it} + \varepsilon_{it} \end{aligned} \tag{4}$$

where $\lambda_{i\tau}$, which can be positive or negative, captures the net improvement in the placement score between attempts $\tau - 1$ and τ. The first term in this equation, $\bar{s}_{it}$, represents the stock of test-taking skills that the student acquires or loses over time. This term evolves over time as student i accumulates more shocks $\lambda_{i\tau}$ under the summation sign in the last line of (4). We call this term a noise-free score. The second term, ε_{it}, captures transitory noise.

Our estimation strategy, outlined in Section 4, relies on the following independence and distributional assumptions that let us identify the model using cross-sectional data only.

[25] While we cannot pinpoint the exact mechanism with total confidence, this correlation suggests that parents with lower income are less willing to invest in their daughters' college education.

Assumption 2. ***(Conditional independence and normality of shocks)*** *In the population of first-time takers, the unobservable components of GPA and subject scores (θ_i, λ_{it}, ε_{ig}, ε_{ijt}) are: (a) independent of X_i and of each other,*[26] *and (b) jointly normal, conditional on i's income.*

As we explain later in Section 4, conditional independence of unobservables plays an important role in the identification argument. Normality is mostly a convenience assumption: it allows us to approximate integrals in the GMM objective function using Gauss-Hermite quadratures; it also simplifies the formal proof of identification.

How strong is our assumption on conditional independence of learning shocks? A concern might be that students who did not study prior to the first attempt may have greater potential for improvement than those who did. In this case, learning shocks λ_{it+1} might depend on g_i and s_{it}. Allowing the distribution of learning shocks to vary across income groups partly addresses this concern: as shown in Table 5, income tightly correlates with pre-exam investment before attempt 1.

Assumption 3. ***(Noise invariance)*** *The distribution of the transitory shocks, ε_{ijt}, does not depend on the attempt, t. It is allowed to depend on subject and student i's income.*

Allowing the distribution of ε_{ijt} to depend on subject is important as the data demonstrates large differences in the score variance across subjects.

Assumption 4. ***(Stationarity after attempt 4)*** *Noise-free scores stop evolving after attempt 4. The same applies to costs of retaking, $\psi_t(I_i)$: $\psi_t(I_i) = \psi_4(I_i)$ for $t > 4$.*

We need this assumption as the number of attempts is censored at 5 in our data. This assumption also agrees with the estimates we present in Section 5. By attempt 4, λ_{it}'s estimated mean and variance fall to zero.

Together, the above assumptions imply

$$\theta_i|I_i \sim N[0, \Sigma_\theta(I_i)], \quad \varepsilon_{it}|I_i \sim N[0, \sigma^2_\varepsilon(I_i)], \quad \varepsilon_{ig}|I_i \sim N[0, \sigma^2_{\varepsilon g}(I_i)],$$
$$\lambda_{it}|I_i \sim N[\mu_{\lambda t}(I_i), \sigma^2_{\lambda t}(I_i)], \quad \lambda_{it} = 0 \text{ and } \psi_t(I_i) = \psi_4(I_i), \forall t > 4.$$

3.2 College Seats and Student Preferences

There is a continuum of seats that correspond to college programs and outside options, such as distance education or dropping out.

Assumption 5. ***(Vertical preferences)*** *All science students rank programs in the same way. While ordinal preferences are the same for everyone, we allow cardinal preferences to depend on income.*

[26] The components of $\theta_i = [\theta_{iq}, \theta_{iv}]$ are allowed to correlate without any restrictions. The same applies to the components of $[\lambda_{i,M,t}, \lambda_{i,T,t}, \lambda_{i,Sc,t}, \lambda_{i,SS,t}]$.

Though this is a strong assumption, it may be less objectionable in the Turkish context as there seems to be a clear hierarchy of programs. The top tier includes world-known public and private universities, while the bottom-tier schools are those offering only two-year programs and distance education. Also, by focusing on the science track only, we hope to reduce the extent of idiosyncratic preferences when choosing a major of study. This assumption provides us with a nice setting to get started as modeling horizontal differentiation along with dynamics complicates things substantially.[27]

While the ranking of programs is the same for everyone, we allow the intensity of preferences to depend on student's income category, I_i. For example, poor students may be more eager to be placed in top programs as they cannot afford high-quality education abroad. Intensity of preferences is also related to attitude towards risk. In effect, allowing the curvature of utility to depend on income class permits differences in risk aversion by income.

Let $r \in [0, 1]$ rank college seats, with $r = 1$ representing the most popular program and $r = 0$ representing an outside option—a seat in a distance education program or dropping out—that is freely available. $U(r; I_i)$ denotes the utility of someone in income group I_i being placed in seat r. $U(r; I_i)$ is non-decreasing in r by construction. We choose measurement units for the placement utility to ensure that $U(0; I_i) = 0$ and $U(1; I_i) = 1$. That is, all welfare statistics we report are relative to the difference in payoffs between the best and the worst possible placement. In what follows, we estimate $U(r; I_i)$ using a flexible approximation that takes values in $[0, 1]$ and is non-decreasing.

3.3 Timing of Events

We assume that student i observes X_i, θ_i, and g_i before taking the exam for the first time. Since the first attempt takes place at the end of high school and almost all high school seniors take the exam, we assume that there are no costs of taking the exam for the first time.

1. Upon receiving his first-time exam score, s_{i1}, the student learns his ε_{i1}. The student also observes the past year's cutoff scores for all seats, $s^*(r)$, which are determined by the actions of the previous cohorts of students.[28] Knowing s_{it}, ε_{it} and the function s^*, all students, first-time and repeat takers alike, simultaneously submit ordered lists of preferred programs to the exam authority.

[27] For example, if a student has a strong personal preference for some unpopular program, he would be more likely to accept placement after the first attempt, compared to his more ambitious peers with the same ability. Thus, decisions to retake are likely to be influenced by student-specific tastes. It is not clear if Hotz-Miller's estimation approach can be adapted to the setting with persistent unobserved heterogeneity in preferences. One could use a nested fixed point estimator similar to Rust (1987) instead, but this would likely be too computationally demanding: on each iteration of the estimator, one would have to solve a separate dynamic problem for each draw of idiosyncratic preferences over hundreds of college programs.

[28] By definition of the cutoff score, a student can be admitted to r if his score is above $s^*(r)$.

2. The exam authority allocates seats based on scores and preference lists using the serial dictatorship mechanism.[29] All placements are terminal.

3. Students who are not placed retake next year.[30] There is a retaking cost, $\psi_t(I_i)$, incurred in attempt t by the student in income group I_i after he decides to retake. This captures the fact that the exam is given only once a year and that preparation is costly. Making costs depend on I_i captures differences in access to tutoring and opportunity costs of waiting an extra year. For instance, low-income students ($i : I_i = 1$) may be pressured to provide for their families instead of preparing for the exam.

4. Second-time takers study for the exam, draw and observe their learning shocks λ_{i2}. Upon receiving their scores in the second attempt, they learn their ε_{i2} and submit their preference lists. A similar time line occurs for later attempts. Future payoffs are discounted at a common rate δ.

We normalize the mass of first time takers to one. Stationarity (assumption 1) requires that the number of high school graduates, the composition of the incoming cohort and the supply of quota-constrained seats are constant over time. As a result, the cutoffs s^* do not change over time.

Thus, exam takers play a game in a stationary overlapping generations environment. The payoffs in this game are given by the placement utility function, retaking costs and the discount factor. Students are uncertain about their future scores, but there is no aggregate uncertainty in the model. A strategy in this game is a mapping that takes student's current state and the cutoff scores as inputs and provides an ordered list of preferences for programs. In the remaining part of this section, we characterize a steady state Markov-perfect equilibrium of this game. In equilibrium, the student wishing to be placed will submit his most preferred program among those feasible given his score. Or, if the student prefers retaking, he submits an empty list.

3.4 Steady State Equilibrium

In what follows, we characterize a steady state of this game, an equilibrium in which: i) the number of placed students is identical to the size of the incoming cohort, and ii) the cutoff scores stay constant over time. In the steady state, students maximize their expected payoffs taking the cutoff scores as given.

Note that equilibrium strategies can be reduced to a simple binary choice. The student can retake in hopes of improving his score by submitting an empty list. Doing so guarantees non-placement, which permits retaking without a penalty. Or he can choose to be placed in the best

[29] The mechanism starts with all seats vacant. It goes from the highest scoring applicant to the lowest scoring one. For each applicant, it finds the top preferred seat from the applicant's list, among those that are still vacant, and places him to this seat. If a student's turn to be assigned arrives when all programs from his list are already taken, he is not placed.

[30] Note that those who do not wish to go to college can drop out, i.e., be placed in program 0.

seat feasible with his current score by submitting a list containing this seat only. This seat's index is given by the inverse of the cutoff score function, $r^*(s_i) = s^{*-1}(s_i)$.[31]

Students maximize their welfare by solving a dynamic optimization problem. Let V_t be the value function for attempt t and VC_t be the expected payoff associated with retaking. Student's state is captured by t, the noise-free score, $\bar{s}_{it}$, and the transitory shock, ε_{it}. Since the payoffs and the costs are allowed to vary across the three income groups, V_t also depends on the income group indicator I_i:

$$\begin{aligned} V_t(\bar{s}_{it}, \varepsilon_{it}, I_i) &= \max\left\{U\left(r^*\left(\bar{s}_{it} + \varepsilon_{it}\right); I_i\right), VC_t\left(\bar{s}_{it}; I_i\right)\right\} \\ VC_t(\bar{s}_{it}, I_i) &= \delta E\left[V_{t+1}\left(\bar{s}_{it} + \lambda_{it+1}, \varepsilon_{it+1}; I_i\right) \middle| \bar{s}_{it}, I_i\right] - \psi_t(I_i). \end{aligned} \tag{5}$$

The expectation in the second line is taken over the next attempt's learning and transitory shocks, λ_{it+1} and ε_{it+1}.

As the placement utility, $U(r; I)$, is non-decreasing in r, and the inverse of the cutoff score, $r^*(s)$, is non-decreasing in s, the payoff from being placed is non-decreasing in ε_{it}. Since the value of retaking, VC_t, does not depend on ε_{it}, while $U\left(r^*\left(\bar{s}_{it} + \varepsilon_{it}\right); I_i\right)$ does, if retaking is optimal for $\varepsilon_{it} = \tilde{\varepsilon}$, it is also optimal for any $\varepsilon_{it} < \tilde{\varepsilon}$. Thus, student's decision should follow a simple threshold rule: retake after the t^{th} attempt if[32]

$$\varepsilon_{it} < e_t(\bar{s}_{it}, I_i). \tag{6}$$

where $e_t(\bar{s}_{it}, I_i)$ denotes the retaking threshold for i. That is, student i has a target score of $\bar{s}_{it} + e_t(\bar{s}_{it}, I_i)$ based on innate ability and accumulated learning shocks. If his actual score is below this target, the student is better off retaking.

Retaking thresholds $e_t(\bar{s}, I)$ for all t, $\bar{s}$ and I are determined by the model's parameters and the inverse placement cutoff function $r^*(s)$. In turn, $r^*(s)$ is fully determined by student retaking behavior and thus by $e_t(\bar{s}, I)$. In Appendix B.2, we lay out the full system of conditions that $e_t(\bar{s}, I)$ and $r^*(s)$ have to satisfy to generate a steady state equilibrium. We also show that given $r^*(s)$, the retaking threshold function $e_t(\bar{s}, I)$ is uniquely defined for every income group I and attempt t.

[31] There is no reason to misrepresent preferences on the part of students. As there is a continuum of students, no single student can affect the cutoffs, which is why these are taken as given. With given cutoffs, there is no strategic element in reporting preferences and so student report truthfully among the feasible options. If seats r and r' are feasible and the student prefers r to r', but lists r' instead of r, he would be placed in r', which is inferior to r. Thus, the student is better off not doing so and reporting truthfully. While students have no incentives to misrepresent their preferences in the model, the mechanism implemented in Turkey may generate such incentives with small numbers. We discuss this issue in Appendix B.1.

[32] Note that we do not require $U(r(\bar{s} + \varepsilon), I)$ to be linear in ε. This contrasts the standard additive separability assumption widely used in the literature on single-agent dynamic models (e.g., see Aguirregabiria and Mira (2010)).

4 Estimation and Identification

Our goal is to estimate the model's structural parameters. We proceed in three steps. First, we estimate the factor loadings and the coefficients from the GPA and score equations (2) and (3): $\beta_g(I_i)$, $\alpha_g(I_i)$, $\beta_j(I_i)$, $\alpha_j(I_i)$, where $I_i = 1, 2, 3,\ j = M, T, Sc, SS$. In this same step, we also recover parameters related to the distributions of innate abilities and idiosyncratic shocks: $\Sigma_\theta(I_i)$, $\sigma_\varepsilon(I_i)$, $\sigma_{\varepsilon g}(I_i)$. In the second step, we turn to the dynamic problem. Our objective here is to find the parameters of the learning shock distributions, $\mu_{\lambda t}(I_i)$ and $\sigma_{\lambda t}(I_i)$, $t = 2, 3, 4$, which determine how student states evolve over attempts. We also estimate non-parametrically the cutoff function $e_t(\overline{s}_{it}, I_i)$, which describes the way students respond to changes in own state. In step 3, we use the above estimates to find payoff-related parameters: the costs of retaking, $\psi_t(I_i)$, and the placement payoff function, $U(r; I_i)$. The latter is identified non-parametrically for all values of $r \in [0, 1]$ and $I_i = 1, 2, 3$ using a flexible approximation. As implied by notation, all of the above elements of the model are identified separately for each of the three income groups.[33]

We discuss the intuition behind each estimation step below. Technical details can be found in the online appendix.

4.1 Step 1: Initial Conditions

Our objective in step 1 is to identify the joint distribution of $(\bar{s}_{i1}, \varepsilon_{i1})$ and g_i among the first-time takers,[34] which is determined by the five-equation system in (2) and (3). Identifying the coefficients on X_i in these equations is straightforward. Since $\theta_i'\alpha_g(I_i) + \varepsilon_{ig}$ and $\theta_i'\alpha_j(I_i) + \varepsilon_{ijt}$ are defined as residuals, the coefficients on X_i come from estimating equations (2) and (3) via OLS in the sub-sample of first-time takers from income groups $I_i = 1, 2, 3$.

One may be concerned that students who have low chances of going to college under the current policy do not take the exam at all. This would lead to sample selection on $\bar{s}_{i1}$. However, as practically every high school senior takes the entrance exam, our sample of first-time takers should be representative of all high school graduates from the science track.

Conditional on the observed covariates, the variation in $(\bar{s}_{i1}, g_i, \varepsilon_{i1})$ is driven by the unobserved ability θ_i and the idiosyncratic shocks. Independence and joint normality of θ_i, ε_{ig} and ε_{ij1} put enough structure on the data to identify the remaining elements of this distribution.[35] Let R_{ig} denote the residual from the GPA equation, $\theta_i'\alpha_g(I_i) + \varepsilon_{ig}$; similarly, let R_{ij} be the residual from the respective subject score equation. The covariance matrix for the vector

[33] For instance, we estimate three scalar parameters $\psi_1(I_i),\ I_i = 1, 2, 3$—the costs of proceeding to attempt 2 faced by low, middle and high-income students, respectively.

[34] Although g_i is not a state variable in the dynamic problem, knowing the joint distribution of $(\bar{s}_{i1}, g_i, \varepsilon_{i1})$ is instrumental in the subsequent steps.

[35] In principle, the densities of θ_i, ε_{ig} and ε_{ij1} could be non-parametrically identified together with the factor loadings as in Bonhomme and Robin (2010) or Freyberger (2012). While the independence assumption is essential, normality is only needed to reduce the computational burden in the later steps.

$R_i = [R_{ig}, R_{iM}, R_{iT}, R_{iSc}, R_{iSS}]'$ is related to the factor loadings, the variances of the idiosyncratic shocks and the covariance matrix of θ_i as follows:

$$Var[R_i|I_i = I] = \alpha'(I)\Sigma_\theta(I)\alpha(I) + \Sigma_\varepsilon(I), \tag{7}$$

where $\alpha(I)$ denotes the matrix of factor loadings of income group I and Σ_ε is a diagonal matrix with the variances of ε_{ig} and ε_{ij1}'s on the main diagonal. Parameters $\alpha(I)$, $\Sigma_\theta(I)$ and $\Sigma_\varepsilon(I)$, which capture the distribution of unobservables, are estimated via GMM using equations in (7) as identifying moment conditions. Knowing the above parameters, as well as $\beta(I)$ and X_i, one can use equations (2) and (3) to find the joint density of $\bar{s}_{i1}$, g_i, and ε_{i1} conditional on X_i.[36]

Note that in the extended version of the model that allows students to choose the high school type, the distribution of $\bar{s}_{i1}$, g_i and ε_{i1} is not a policy-invariant object. For instance, a restriction on retaking may force more high school students to use private tutors. This would change the distribution of X_i in the population of first-time takers, which would in turn affect the distribution of $\bar{s}_{i1}$. Whenever we use the extended model, we explicitly account for such changes in X_i as described in Appendix A.3.

4.2 Step 2: State Transitions and Students' Policy Function

Our method of identifying the dynamic component of the model uses insights from the conditional choice probability approach, first proposed by Hotz and Miller (1993).[37] Unfortunately, we cannot apply this approach off-the-shelf. Unlike the original setting in Hotz and Miller (1993), our model and data do not permit straightforward estimation of state transition probabilities or policy functions. For instance, in Hotz-Miller's case, student's noise-free scores in attempts 1 and 2, $\bar{s}_{i1}$ and $\bar{s}_{i2}$, would be observable. An empirical distribution of $\bar{s}_{i2} - \bar{s}_{i1}$ would be a consistent estimator for the distribution of λ_{i2}, the learning shock between attempts 1 and 2. In our case, however, the following complications arise: i) we observe $s_{i1} = \bar{s}_{i1} + \varepsilon_{i1}$ and $s_{i2} = \bar{s}_{i2} + \varepsilon_{i2}$, so that $\bar{s}_{i2} - \bar{s}_{i1}$ is observed with noise, ii) the noise is not likely to be mean-zero as selection into retaking is based on

---

[36] The loadings in the math and Turkish equations are normalized, respectively, to $\alpha_M = [1, 0]'$ and $\alpha_T = [0, 1]'$: in other words, quantitative ability is a common factor that affects the math score but not the Turkish score and vice versa for verbal ability.

This normalization is without loss of generality. Suppose, for example, that $\alpha_M \neq [1, 0]'$. One can redefine common factors as follows: $\widetilde{\theta}_{iM} = \theta_i'\alpha_M$ and $\widetilde{\theta}_{iT} = \theta_i'\alpha_T$. Rewritten in terms of $\widetilde{\theta}_i$, the math score equation from the system (3) will include $\widetilde{\theta}_{iq}$ with a coefficient of one and will not include $\widetilde{\theta}_{iv}$ at all. Thus, the original system of equations in (3) can always be rewritten to ensure that the factor loadings in the math equation are equal to $[1, 0]$. What we identify as factor loadings in the empirical exercise are the coefficients on $\widetilde{\theta}_i$ in the rewritten system; what we identify as a covariance matrix Σ_θ is the covariance matrix of $\widetilde{\theta}_i$. Appendix B.4 provides more detailed derivations on this point.

[37] One rationale for using the conditional choice probability approach is that it does not require equilibrium uniqueness. CCP is very popular, for instance, in the literature on estimating dynamic games where the issue of equilibrium multiplicity is central (Bajari, Benkard, and Levin (2007) is one such example). Using the CCP approach relieves us from proving that the equilibrium in our model is unique.

ε_{i1}, iii) we do not have panel data to compute improvement in scores at the student level, $(s_{i2} - s_{i1})$. We can compare the exam score distributions of first-time takers to those of repeat takers, but in the absence of extra information, one cannot disentangle the differences due to learning from those due to selection.

Given our data constraints, we use a novel approach that relies on the fact that high school GPA is not affected by retaking, which implies that the GPA distribution of repeat takers differs from that of first-time takers only because of selection. The distribution of exam scores of repeat takers, in contrast, is affected by both selection and learning. Thus, by comparing the distributions of scores and GPAs across attempts, we are able to distinguish learning from selection. We assume steady state so that second-time takers in a given year can be thought of as identical to retakers from today's cohort of first-time takers and so on.

Below we depict how selection and learning operate in a very simple example that highlights the essential intuition.[38] We use first and second-time takers for concreteness. Assume there are two types, labeled high and low, with half of the agents being of each type. GPA and placement scores are

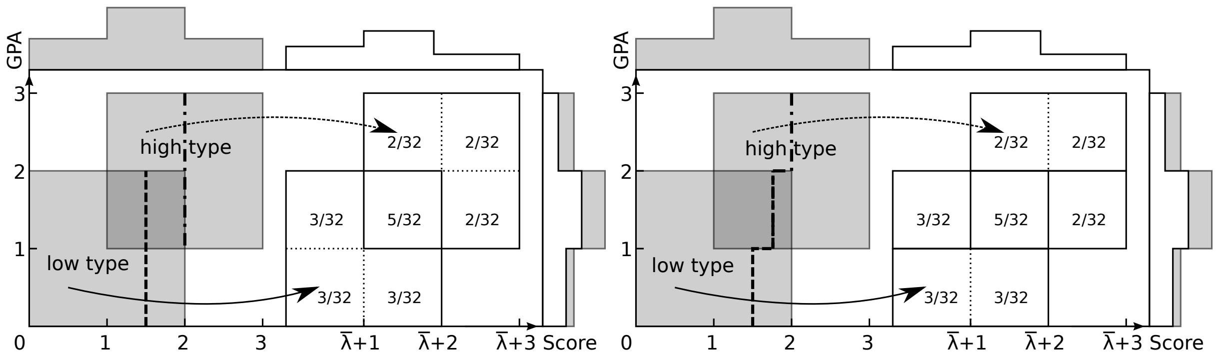
$$\begin{aligned} g_i &= \theta_i + \varepsilon_{ig} \\ s_{i1} &= \theta_i + \varepsilon_{i1} \\ s_{i2} &= \theta_i + \lambda_{i2} + \varepsilon_{i2} \end{aligned}$$

where $\theta_i = 2$ for i if he is a high type and $\theta_i = 1$ if he is a low type. The shocks ε_{ig}, ε_{i1}, ε_{i2} are drawn from a uniform distribution over $[-1, 1]$. Learning shocks are deterministic and the same for both types, so $\lambda_{i2} = \bar{\lambda}$. Note that this example differs from the estimated model as it assumes placement scores do not include GPA as a component (i.e., $w_{ig} = 0$). As a result, retaking cutoffs vary only by ability type. In contrast, in our estimated model, retaking cutoffs depend on the permanent part of the score (i.e., type), which includes the GPA among other components of the score.

The joint distribution of g_i and s_{i1} for the universe of agents is depicted in Figure 2a, where we have the high school GPA and the placement score on the two axes. The shaded lower left square corresponds to the low type, while the shaded upper right square depicts the high type. The retaking rule is to retake if $\varepsilon_{i1} < e_1(\bar{s}_{i1})$ and hence if $s_{i1} < \bar{s}_{i1} + e_1(\bar{s}_{i1})$. In this simple model, $\bar{s}_{i1}$ is just θ_i so that the two types have different cutoffs as depicted by the two vertical dashed lines. We have low types being more likely to retake in this figure. The shaded histograms on the right and

[38] A more formal identification argument can be found in Appendix B.5.

Figure 2. Identifying Learning and Selection Using Scores and High School GPA



(a) Base line: Learning shocks do not depend on GPA

(b) Learning shocks depend on GPA

Notes: The boxes inside each diagram depict the joint distributions of scores and GPAs in attempts 1 and 2. The histograms at the top are marginal distributions of scores (inflated by the number of students). The histograms on the right hand side of each diagram show marginal distributions of GPA. Distributions related to the first attempt are shaded, while distribution of second time takers are in white. The dashed lines depict the retaking thresholds, $\bar{s}_{it} + e_1(\bar{s}_{it})$, for each student type.

the top of the figure denote the marginal distribution of GPA and scores among first-time takers, respectively.

The unshaded histograms depict the distributions of GPA and scores, joint and marginal, of second-time takers. Half the high type and three-quarters of the low type of agents retake. This results in the mass in the upper tail of the GPA distribution of second time takers being lower than that in the lower tail. The learning shocks are depicted by the arrows. Given our assumptions on learning shocks, the distribution of scores moves to the right by $\bar{\lambda}$. The number in each unshaded box is equal to the mass of second-time takers there. The total mass equals the share of first-time takers who choose to retake.

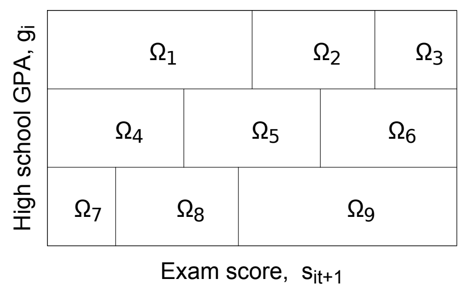
We identify the cutoff scores for the two types by comparing the distributions of GPAs of first and second-time takers. The mass at any point in the upper tail of the GPA distribution of second-time takers gives us the cutoff for the high type, while the mass at any point in the lower tail pins down the cutoff for the low type. The shift in the score distribution between the first and second attempt pins down the learning shock. If learning shocks are stochastic, in addition to being shifted to the right as depicted in Figure 2a, the distribution of scores will gain more variance due to variation in λ_{i2}. This extra variance allows us to identify $\sigma^2_{\lambda 2}(I)$, the variance of learning shocks in attempt 2. A similar argument applies to third versus second-time takers, and so on.

Note that the above argument relies on the assumption that learning shocks are independent of ability. In Figure 2b, we modify the above example to allow the distribution of learning shocks to depend on GPA. This constructed example is observationally equivalent to the one depicted in Figure 2a. That is, the set of cutoffs and learning shocks depicted in Figure 2a gives the same GPA and score distributions for second-time takers as that in Figure 2b. Learning shocks are $\bar{\lambda}$ for both types if their GPAs are in the non-overlap region (below 0.5 and above 1). In the overlap region, the learning shocks are stochastic: high types get $\lambda_{i2} = \bar{\lambda}$ with probability 0.4 and $\lambda_{i2} = \bar{\lambda} - 1$ otherwise. Low types get $\lambda_{i2} = \bar{\lambda}$ with probability 0.6 and $\lambda_{i2} = \bar{\lambda} + 1$ otherwise. As the future score distribution looks the same for the high and low types (both have mean ability $(\bar{\lambda} + 1)$) 60 percent of the time and $(\bar{\lambda} + 2)$ otherwise) their cutoff is the same as depicted in Figure 2b.

The score and GPA distributions of second-time takers are exactly the same as those in the previous example. Thus they are observationally equivalent, and the model cannot be identified.

4.2.1 Estimation algorithm for attempts 1–3.

Figure 3. Sets of GPAs and Scores Used to Construct the GMM Estimator



The retaking threshold function and the parameters of learning shocks, $e_t(\bar{s}; I)$, $\mu_{\lambda t+1}(I)$ and $\sigma_{\lambda t+1}(I)$, are estimated sequentially using a separate GMM routine for each attempt $t = 1, 2, 3$ and income group $I = 1, 2, 3$. This GMM estimator is designed to match predicted numbers of exam takers whose GPAs and scores fall into sets $\Omega_1, \ldots, \Omega_9$, depicted in Figure 3. The horizontal lines in this figure are the first and second terciles of the GPA distribution, while the vertical lines are exam score terciles conditional on GPA being between the respective horizontal lines.

Let $\bar{S}_{it} = [\bar{s}_{i1} \ldots \bar{s}_{it}]$ denote the trajectory of noise-free scores that student i would face if he takes the exam at least t times, and $F(\bar{S}_{it}|s_{it}, g_i, X_i, I_i)$ be the probability measure in the space of such trajectories conditional on the observables in attempt t.[39] Let a_{it} be a dummy that equals one if i is not placed after t attempts. The set of moment conditions that are used to estimate

[39] Note that $\bar{S}_{it}$ is exogenous; by assumption 2, the student cannot affect learning shocks λ_{it}, which drive the evolution of $\bar{s}_{it}$.

$e_t(\cdot, I)$, $\mu_{\lambda t}(I)$ and $\sigma_{\lambda t}(I)$ take the following form:[40]

$$\Pr[(g_i, s_{it+1}) \in \Omega_k, a_{it+1} = 1 | a_{it} = 1, I_i = I] = \int \left[\int \Pr\{a_{it} = 1 | \bar{S}_{it}, I\} \Phi\left(\frac{e_t(\bar{s}_{it}, I)}{\sigma_\varepsilon(I)}\right) \times \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}; \mu_{\lambda t+1}(I), \sigma_{\lambda t+1}(I)\} dF(\bar{S}_{it} | s_{it}, g_i, X_i, I) \right] \frac{dG(s_{it}, g_i, X_i | I, a_{it} = 1)}{\Pr\{a_{it} = 1 | s_{it}, g_i, X_i, I\}} \quad k = 1, \ldots, 9. \quad (8)$$

where G denotes the distribution of the observables among t-time takers from the income group I.

The expression on the right hand side of (8) predicts the share of applicants who retakes after attempt t and whose score-GPA combination in attempt $t+1$ ends up in the cell Ω_k. We take an expectation over all observables among t-time takers, s_{it}, g_i and X_i, and trajectories $\bar{S}_{it}$ that the student's noise-free scores can follow conditional on the observables. The first term inside the inner integral is the probability that the trajectory is not interrupted by a placement decision before attempt t. The second is the probability that the student retakes at t. Finally, the third term gives the probability that the trajectory of scores lands into Ω_k in attempt $t+1$. The inner integral is divided by the probability of surviving t attempts conditional on attempt t's observables:

$$\Pr\{a_{it} = 1 | s_{it}, g_i, X_i, I\} = \int \Pr\{a_{it} = 1 | \bar{S}_{it}, I\} dF(\bar{S}_{it} | s_{it}, g_i, X_i, I) \quad (9)$$

In order to use conditions (8) in a GMM estimator, one has to compute the integrands in (8) and (9) and approximate the expected values with the finite-sample analogs.

First, the probability of surviving t attempts along the trajectory $\bar{S}_{it}$ is easy to find from (6). The student keeps retaking if his ε shocks received in all prior attempts stay below the respective retaking thresholds. Therefore, the above probability can be approximated by

$$\widehat{\Pr}\{a_{it} = 1 | \bar{S}_{it}, I\} = \prod_{\tau=1}^{t-1} \Phi\left(\frac{\widehat{e}_\tau(\bar{s}_{i\tau}, I)}{\hat{\sigma}_\varepsilon(I)}\right) \quad (10)$$

The probability that the combined shock $\lambda_{it+1} + \varepsilon_{it+1}$ takes the student from $\bar{s}_{it}$ into Ω_k is approximated by

$$\widehat{\Pr}\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}, I; \mu_{\lambda t+1}, \sigma_{\lambda t+1}\} = \begin{cases} 0, \text{if } (g_i, s) \notin \Omega_k \ \forall s, \\ \Phi\left(\frac{S_{ku} - \bar{s}_{it} - \mu_{\lambda t+1}}{\sqrt{\hat{\sigma}_\varepsilon^2(I) + \sigma_{\lambda t+1}^2}}\right) - \Phi\left(\frac{S_{kl} - \bar{s}_{it} - \mu_{\lambda t+1}}{\sqrt{\hat{\sigma}_\varepsilon^2(I) + \sigma_{\lambda t+1}^2}}\right), \text{ otherwise,} \end{cases} \quad (11)$$

[40] Appendix B.6 formally derives these moment conditions.

where S_{ku} and S_{kl} are the upper and the lower boundaries of scores in Ω_k.

Integrating over $\bar{S}_{it}$ in (8) and (9) requires the knowledge of joint density of $\bar{s}_{i1}, \ldots, \bar{s}_{it}$ conditional on s_{it}, g_i, X_i and I_i. Note that this density describes the whole population, not just the retakers who survive t attempts. The latter fact allows us to use normality and independence of all shocks in the model. Given assumptions 2 and 3, the variables $\bar{s}_{i1}, \ldots, \bar{s}_{it}, s_{it}$, g_i are jointly normal conditional on X_i and I_i in the original cohort of students. The mean and the covariance matrix of this distribution depend on $\{\mu_{\lambda\tau}(I_i), \sigma_{\lambda\tau}(I_i)\}_{\tau=2}^{t}$ and the parameters estimated in step 1. This implies that the sequence $\bar{s}_{i1}, \ldots, \bar{s}_{it}$ is jointly normal, too, conditional on s_{it}, g_i, X_i and I_i. The mean and the covariance matrix of this distribution can be easily derived from the above parameters.

In order to obtain finite-sample analogs for the moment conditions in (8), we approximate the left-hand side by the number of $(t+1)$-time takers in the set Ω_k divided by the total number of t-time takers in the data. The outer integral on the right-hand side is approximated by the average over (s_{it}, g_i, X_i) of t-time takers in the data who belong to income group I. The inner integral is computed numerically using Gauss-Hermite quadratures.

We approximate the unknown retaking thresholds $e_t(\bar{s}, I)$ by piecewise-linear functions defined on three grid points, $s_{1,It}, s_{2,It}, s_{3,It}$:

$$e_t(\bar{s}, I) = \begin{cases} e_{1,It}, & \bar{s} \leq s_{1,It} \\ e_{1,It} + (e_{2,It} - e_{1,It})\frac{\bar{s}-s_{1,It}}{s_{2,It}-s_{1,It}}, & s_{1,It} \leq \bar{s} \leq s_{2,It} \\ e_{2,It} + (e_{3,It} - e_{2,It})\frac{\bar{s}-s_{2,It}}{s_{3,It}-s_{2,It}}, & s_{2,It} \leq \bar{s} \leq s_{3,It} \\ e_{3,It}, & s_{3,It} \leq \bar{s} \end{cases}$$

The grid points $s_{1,It}$ and $s_{3,It}$ are located at the 20^{th} and the 80^{th} percentiles of s_{it} among t-time takers from the income group I, while $s_{2,It} = (s_{1,It} + s_{3,It})/2$. In total, we have nine equations in (8) to identify five parameters, $e_{1,It}$, $e_{2,It}$, $e_{3,It}$, $\mu_{\lambda t+1}(I)$, $\sigma_{\lambda t+1}(I)$, for each combination of I and t.

4.2.2 Estimation for Attempts Greater Than 3

As we mentioned in Section 3, the number of attempts is censored at 5 in our data. Thus, we cannot use moment conditions (8) for $t \geq 4$. Instead, we exploit our assumption 4 to derive a separate set of identifying equations for all such attempts. We rely on one important implication of this assumption: student's future stream of payoffs does not depend on t after $t = 4$. Thus, $e_4(\bar{s}, I) = e_t(\bar{s}, I)$ for any $t \geq 4$. In order to completely describe student's behavior after fourth attempt, one has to pin down a single threshold function $e_4(\bar{s}, I)$.

The set of moment conditions for $e_4(\cdot, I)$ is obtained by summing both sides of equation (8) for attempts $t \geq 4$ and expressing the right-hand side in terms of variables observed in attempt

4 (see Appendix B.6):

$$\sum_{t\geq 4} \Pr[(g_i, s_{it+1}) \in \Omega_k, a_{it+1} = 1 | a_{i4} = 1, I_i = I] = \int \left[\int \Pr\{a_{i4} = 1 | \bar{S}_{i4}, I\} \frac{\Phi\left(\frac{e_4(\bar{s}_{i4}, I)}{\sigma_\varepsilon(I)}\right)}{1 - \Phi\left(\frac{e_4(\bar{s}_{i4}, I)}{\sigma_\varepsilon(I)}\right)} \right.$$

$$\left. \times \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{i4}, I\} dF(\bar{S}_{i4} | s_{i4}, g_i, X_i, I) \right] \frac{dG(s_{i4}, g_i, X_i | I, a_{i4} = 1)}{\Pr\{a_{i4} = 1 | s_{i4}, g_i, X_i, I\}}$$

$$k = 1, \ldots, 9. \quad (12)$$

In principle, one could estimate the full set of parameters for all attempts and income groups in one run. However, we choose to run a sequence of GMM estimators to avoid computational issues. First, we estimate parameters associated with the first retaking decision: $e_1(\cdot, I)$, $\mu_{\lambda 2}(I)$ and $\sigma_{\lambda 2}(I)$. Then, we set up and run the GMM estimator for $e_2(\cdot, I)$, $\mu_{\lambda 3}(I)$ and $\sigma_{\lambda 3}(I)$, using $\widehat{e}_1(\cdot, I)$, $\widehat{\mu}_{\lambda 2}(I)$ and $\widehat{\sigma}_{\lambda 2}(I)$ to compute the probability of survival in (10) and the distribution of noise-free scores F. Then, we obtain the estimates for $t = 3$ in a similar way; we use the GMM estimator based on (12) for $t = 4$.[41]

4.3 Step 3: Payoff-Related Parameters

In the final step of our estimation procedure, we find the fundamental components of the model related to payoffs: the costs of retaking, $\psi_t(I_i)$, and the utility function, $U(r, I_i)$. We do so along the lines of step 2 in Hotz-Miller's CCP algorithm.

First, note that the continuation value function $VC_t(\bar{s}_{it}, I_i)$ from Bellman's equation (5) can be found by integrating the net present value of future payoffs over all trajectories $\bar{s}_{it+1}, \bar{s}_{it+2}, \ldots$ and all future student actions. Since we have estimated the parameters of learning shocks in step 2, we know the law that generates state transitions: $(\bar{s}_{i\tau+1} - \bar{s}_{i\tau}) \sim N[\mu_{\lambda\tau+1}(I_i), \sigma_{\lambda\tau+1}(I_i)]$, $\varepsilon_{i\tau} \sim N[0, \sigma_\varepsilon(I_i)]$. We also know the decision rule that students use in equilibrium: retake in attempt τ if $\varepsilon_{i\tau} < e_\tau(\bar{s}_{i\tau}, I_i)$. Thus, given candidate values of $\psi_t(I_i)$ and $U(r, I_i)$, we can compute the continuation value $VC_t(\bar{s}_t, I)$ on a grid of $\bar{s}_t$ for every income group I by simulating shocks that hit students, student reactions to these shocks and the associated payoffs. Knowing who chooses placement and what their scores are, we can also construct the inverse cutoff function, $r^*(s)$.

Once the continuation values are known for every income group, attempt and grid point, one can impose the assumption that students maximize their utility. That is, a rational student in

[41] Estimating parameters sequentially is less efficient than using one GMM routine with optimal weights. However, a joint GMM routine requires computing the probability in (9), which is very demanding computationally. If we were estimating all parameters in one run, we would have to recompute this probability on every iteration of the GMM algorithm.

the state $(\bar{s}_{it}, \varepsilon_{it})$ retakes if the utility of placement is lower than the continuation value:

$$U(r^*(\bar{s}_{it} + \varepsilon_{it}), I_i) < VC_t(\bar{s}_{it}; I_i).$$

We use this inequality to find the threshold for ε_t, $\tilde{e}_t(\bar{s}_{it}, I_i)$, below which a rational student chooses to retake.

Finally, we plug the threshold functions $\tilde{e}_t(\bar{s}_{it}, I_i)$ into moment conditions (8) and (12) in place of $e_t(\bar{s}_{it}, I_i)$. We find the estimates for $\psi_t(I_i)$ and $U(r, I_i)$ by minimizing the objective function for unweighted GMM employing all moment equations in (8) and (12). In contrast to step 2, we use all moment conditions simultaneously since one of the parameters being estimated (the utility of placement) is common to all attempts.

The above paragraph highlights the main difference between our step 3 and step 2 in Hotz and Miller (1993), upon which our approach builds. A direct application of Hotz-Miller's approach would match the probability of $\varepsilon_{it} < \tilde{e}_t(\bar{s}_{it}, I_i)$ to the observed retaking choices of students. Since we do not observe student choices in the current attempt, we have to indirectly back them out from the distribution of scores and GPAs in attempt $t + 1$.

The utility function is parametrized in a flexible manner as:

$$U(r^*(s), I) = \sum_{j=1}^{10} \gamma_j(I) \Phi\left(\frac{s - s_j}{h}\right), \quad \gamma_j(I) \geq 0, \quad \sum_{j=1}^{10} \gamma_j(I) = 1.$$

The coefficients γ are allowed to differ by income group. The normalization of $\sum_{j=1}^{10} \gamma_j(I) = 1$ ensures that the utility at $s = \infty$ is unity. As $\Phi(.)$ is increasing in s, constraining $\gamma_j \geq 0$ ensures that the utility function is non-decreasing. The larger is h, the smoother is the function; we set $h = 15$.

Note that γ is not a structural parameter; the function approximated by γ depends on $r^*(\cdot)$, an equilibrium outcome, which changes in response to policy interventions. After obtaining the estimates of γ, we use the simulated $r^*(\cdot)$ to find $U(r, I)$ for the values of $r = 0, 0.01, \ldots, 1$. In total, we have four cost and 10 utility parameters for each income group: $\psi_1(I) \ldots \psi_4(I)$, $\gamma_1(I) \ldots \gamma_{10}(I)$. They are identified using 36 moment conditions per income group in equations (8) and (12): nine conditions (one for each cell Ω_k) for each attempt $t = 1, 2, 3$ in (8) and nine more in (12).

The economics behind identification of the payoff parameters is as follows: if the marginal utility of a higher score increases sharply at s, then students close to s will be risk loving and hence tend to retake the exam more than students with s where utility of the score is less convex. Thus, the observed local retaking rates pin down the curvature of the utility function. The retaking costs are pinned down by the overall retaking rates in the given attempt. We do not attempt to estimate

δ, as it is well known that discount factors are hard to identify in such settings (see Magnac and Thesmar (2002)); we set δ at 0.9 for all students.

Overall, our approach to estimating the dynamic model in (5) has many parallels with the conditional choice probability method originally proposed in Hotz and Miller (1993). Similar to Hotz and Miller, we start by identifying agents' optimal decision rule in a flexible way (retaking cutoffs $e_t(\bar{s}_{it}, I_i)$ in our context) and the law that governs state transitions (distributions of ε_{it} and λ_{it}). We use Hotz-Miller's representation to express the value function in (5) in terms of the estimated decision rule, the transition probabilities and the payoff fundamentals, $\psi_t(I_i)$ and $U(r; I_i)$. Then, we find $\psi_t(I_i)$ and $U(r; I_i)$ that minimize the distance between the actual and the predicted retaking rates within subgroups of students defined by their GPAs and exam scores.

The model we estimate, however, departs from the standard Hotz-Miller setting in two important ways. First, the persistent component of the student state, $\bar{s}_{it}$, is not observed in the data; instead, we observe its noisy measure $s_{it} = \bar{s}_{it} + \varepsilon_{it}$. This issue is further compounded with endogenous sample selection based on the measurement error, ε_{it}. Second, we do not have panel data on individual students. For instance, if we observe a repeat taker, we only know that he chose retaking previously; we do not know his past scores or whether he retakes after the current attempt. These features of our model prevent us from applying Hotz-Miller's estimator off the shelf. In particular, estimating state transition probabilities and policy functions is far from straightforward in our case.

5 Estimation Results

In what follows, we present the estimates from the baseline version of the model. Since our estimation algorithm consists of multiple steps and each step treats previously estimated parameters as given, we use bootstrap simulations to compute all reported standard errors and confidence intervals.[42]

5.1 Explaining Variation in Scores and High School GPA

In step 1, we estimate how observables are related to performance measures: GPA in equation (2) and the four subject exam scores in the system of equations (3). We also estimate the variance of the transient shocks to these scores, $\sigma_\varepsilon^2(I)$, the factor loadings, $\alpha_g(I)$ and $\alpha_j(I)$, and the covariance matrix of unobserved ability, $\Sigma_\theta(I)$. We obtain these estimates for each income group I and report them in Tables 7 and 8 below.

Recall that we have normalized scores so that the loading on quantitative (verbal) ability is zero (one) for the Turkish score. Similarly, quantitative (verbal) ability has a unity (zero) weight in

[42] We run all three steps of the estimation algorithm, perform the counterfactual experiments and save the results, repeating this for 15,000 random bootstrap samples. We use the saved results to construct confidence bounds for each estimated parameter and each outcome variable in our policy simulation exercises.

the Math score equation. Our results indicate that the loading on the quantitative portion of ability for the science score is higher than for the social studies and vice versa for verbal ability. We also find that the covariance between θ_{iq} and θ_{iv} is positive and significant, implying that students who are good at math also tend to be good at Turkish.

It is worth noting that, while factors such as parents' occupation, income, and education do seem to positively correlate with performance measures, the size of the coefficients tends to be small. In general, they explain about 10 percent of a standard deviation.[43] In contrast, the coefficients on prep school expenditure and school type are much higher.

The student's gender also seems to be associated with performance: women do better while in high school but men catch up and surpass them in all subjects but Turkish in the entrance exam. In Turkish, women score higher, and this effect is quite large, roughly one third of a standard deviation. These results are stable across income groups: when we split the sample by income and run the regressions separately for each income group, we find the same patterns.

Table 6 summarizes the explained variance that comes from observables, unobserved ability, and noise. These numbers are relatively stable across income groups, with noise contributing only about 6 percent and observables and unobserved ability being roughly equally important. The low contribution of noise suggests that retaking in response to these shocks plays a limited role.

Income	Observables, X	Unobserved ability, θ	Noise, ε_1
Low	49%	45%	6%
Middle	47%	47%	6%
High	43%	51%	6%

Table 6. Contribution of Observables and Shocks

5.2 Improvements in Scores Between Attempts

In step 2, we estimate the distribution of learning shocks. The estimates for expected improvements in placement scores between attempts are reported in Table 9.

Students from high-income families lose ground to the other two groups; students in the middle income group improve the most. This is roughly in line with Frisancho, Krishna, Lychagin, and Yavas (2016), which used a different approach to deal with selection. Marginal learning (of about 8 points) is largest in the second attempt. It is quite large in magnitude as it is roughly a third of a standard deviation of the first-time score. Marginal learning falls to 1.58 and 1.05 in the third and fourth attempt. To place these numbers into context, we can compare them to the difference in average scores in private schools versus public schools. Cumulative learning after three retakes

[43] This suggests that the raw correlation between income and performance often seen in the data is being captured by our other controls.

is roughly of the same magnitude as the effect of graduating from a private school. Differences in learning gains across income groups suggests that a ban on retaking would have distributional consequences; as the rich learn less when retaking, they are the ones who lose the least from the ban.

We also find that there is considerable variance in the learning shocks, as shown in Table 9, which suggests that students are prone to retake as learning becomes a lottery. Students who are close to getting into a prestigious college may thus retake on the off chance of getting in. The variance of learning shocks tends to be positively associated with income, although

Income	Low					Middle					High				
Outcome variable	GPA/2	M	S	SS	T	GPA/2	M	S	SS	T	GPA/2	M	S	SS	T
Father's occupation (base category – employer)															
Works for wages/salary	-0.55	0.10	0.51	1.08	0.25	1.11**	2.02**	2.14**	0.92	-0.48	0.80**	1.20*	1.65**	1.74**	1.09*
Self-employed	-0.58	0.30	0.50	1.06	0.35	1.08**	2.20**	1.93**	0.10	-0.97	0.29	0.23	1.08*	1.64**	1.42**
Unemployed/not in LF	-0.74	0.16	0.52	1.16	0.14	0.87	1.96*	1.94*	0.55	-0.87	0.20	-0.25	0.78	2.90**	0.91
Mother's occupation (base category – employer)															
Works for wages/salary	3.88**	0.73	2.53	1.53	-0.22	-1.32	-4.38	-2.59	-4.05	1.09	1.46*	2.31	1.94	1.92	2.33
Self-employed	2.47**	-1.19	0.57	0.24	-2.74	-1.17	-4.06	-1.64	-3.02	0.54	1.96**	3.81**	2.84*	1.06	1.40
Unemployed/not in LF	3.11**	0.26	1.55	0.56	-1.57	-1.02	-3.66	-1.81	-3.07	1.56	1.92**	2.72*	1.94	1.89	3.00*
Father's education (base category – primary school or lower level)															
Middle/High school	-0.11	0.09	0.02	0.49	0.23	0.27	0.63	0.41	0.13	-0.21	-0.45	-1.06	-0.11	-0.67	0.39
2-year higher education	-0.15	0.25	1.11	0.82	1.65	0.69**	1.65**	1.21*	0.15	0.43	0.25	0.10	1.21	0.06	1.42*
College/Master/Phd	0.88	1.85	2.06*	2.83*	2.34*	0.98**	2.04**	2.17**	1.24**	1.46**	0.61*	0.99	2.04**	1.22	2.18**
Missing	-0.37	-0.72	-0.09	-1.32*	-1.36	0.26	0.79	0.56	0.54	-0.83	-0.60	-2.74*	-1.38	-0.65	-0.08
Mother's education (base category – primary school or lower level)															
Middle/High school	-0.37	-0.12	-0.47	-0.15	0.82	-0.36*	-0.88**	-0.51	-0.38	0.32	-0.44*	-0.56	-0.99*	0.37	-0.21
2-year higher education	-1.32	0.26	-3.54	0.27	3.36	0.13	-0.14	0.50	0.67	2.18**	0.60	0.88	0.50	0.98	1.70**
College/Master/Phd	-2.08	-3.87	-1.00	-0.79	0.21	-0.14	-1.31	-0.49	0.40	0.40	0.81**	1.03	1.13	2.05**	1.94**
Missing	0.53	0.68	-0.38	1.17	1.68*	-0.47	-0.94	-0.43	-1.33	1.06	0.68	2.55	1.66	0.91	1.01
Likely source of income in college (base category – family, rental income or scholarship)															
Work	-0.55**	-1.34**	-1.25**	-0.64	-0.96**	-0.47**	-0.43	-0.96**	-0.78*	-0.88**	-0.69**	-1.06**	-0.80*	0.02	-0.30
Loan	-0.03	0.36	0.20	-0.02	-0.01	0.18	0.69*	0.32	-0.77**	-0.51	0.11	0.78*	0.39	-0.70	-0.68*
Other	-0.60	-0.51	-0.39	0.08	-0.54	-0.75*	-0.64	-1.38**	-0.67	-0.59	-0.77*	-0.13	-1.58*	-1.83*	-0.79

Column titles: GPA – high school GPA, M – math score, S – science score, SS – social sciences score, T – Turkish score. Significance levels: * – 5%, ** – 1%. Standard errors are obtained by bootstrapping (15,000 bootstrap samples used).

Table 7. Estimates of the GPA and the Score Equations (Continued on the Next Page)

Income	Low					Middle					High				
Outcome variable	GPA/2	M	S	SS	T	GPA/2	M	S	SS	T	GPA/2	M	S	SS	T
Internet access (base category – at home)															
Yes, not at home	0.56	1.60	0.49	-0.26	0.38	0.06	-0.02	0.49	-0.63	-0.33	-0.40*	-0.74*	-0.71*	-0.94*	-0.88**
No	0.86	2.14*	0.26	-1.59	-0.56	0.48*	0.64	0.47	-1.28**	-0.45	-0.58**	-1.31**	-1.67**	-2.00**	-1.75**
Missing	0.15	1.48	-0.53	-2.63*	-1.63	0.28	-0.39	0.07	-1.24	-1.50*	-1.90**	-3.74**	-3.53**	-2.73**	-3.45**
Population in the city where the high school is located (base category – less than 10,000)															
10-50k	-0.69*	0.20	-0.70	-1.00*	-0.60	-0.76*	1.77**	0.97	0.05	0.23	-0.40	0.33	0.25	-1.29	-0.92
50-250k	-0.52	1.67**	0.91	1.06*	1.76**	-0.23	4.16**	3.19**	2.36**	1.93**	0.15	2.02	1.62	0.58	1.14
250-1000k	-0.35	2.08**	1.19*	1.04*	1.28*	-0.18	3.85**	2.91**	2.25**	2.61**	0.08	2.37*	1.93*	0.82	1.36
>1 million	-0.19	3.15**	1.81**	2.32**	2.36**	0.31	5.01**	3.68**	3.62**	3.94**	0.80	3.80**	3.51**	3.61**	3.24**
Missing	-0.99**	0.37	-0.54	-0.49	-0.34	-1.04**	1.91**	1.58*	1.52*	1.43*	-0.07	2.14	1.10	1.07	1.30
Prep. school expenditures (base – did not attend prep. schools)															
Scholarship	5.27**	12.2**	12.2**	7.81**	7.46**	6.77**	15.2**	16.2**	11.6**	11.1**	6.32**	14.0**	16.3**	12.9**	11.1**
Less than 1b	2.15**	7.75**	6.27**	1.64**	3.09**	2.27**	7.73**	6.84**	1.77**	3.15**	1.74**	6.89**	5.95**	0.44	2.43**
1-2 b	2.14**	8.13**	5.89**	1.45*	3.47**	1.87**	7.77**	6.16**	0.79	2.58**	0.41	4.70**	3.49**	-1.65*	1.09
More than 2b	1.10	5.46**	5.59**	1.60	4.01**	1.71**	7.48**	6.68**	1.65*	3.06**	0.68	5.62**	3.98**	-1.14	2.14**
Missing	-0.37	-0.05	0.25	0.90**	0.31	0.04	0.77	0.97*	0.39	0.05	-0.48	-1.13	-0.43	-0.72	-0.08
High school type (base – public school)															
Private	2.77**	7.22**	6.40**	2.63**	4.59**	2.47**	6.88**	5.24**	3.04**	5.30**	3.18**	6.46**	5.30**	3.90**	6.02**
Anatolian/Science	5.64**	13.0**	12.9**	7.99**	9.25**	6.20**	13.6**	13.1**	9.90**	10.9**	6.75**	13.8**	13.6**	9.85**	10.5**
Other	-0.29	-1.20	-0.83	1.18*	0.52	0.51	1.50*	1.32*	3.07**	3.72**	0.79	1.68	1.91	4.51**	5.31**
Male	-0.68**	4.54**	3.85**	1.90**	-4.85**	-1.39**	3.02**	2.91**	2.24**	-4.88**	-1.82**	2.18**	2.15**	2.49**	-4.19**
>3 kids in the family	0.02	-0.58	-0.20	-0.64*	-0.77**	-0.03	-0.16	-0.31	-0.54	-1.03**	-0.33	-0.59	-0.31	-0.48	-1.20**
Factor loadings															
on θ_1 (quant. ability)	0.31**	1.00**	0.70**	-0.14**	0.00	0.32**	1.00**	0.75**	-0.28**	0.00	0.32**	1.00**	0.80**	-0.48**	0.00
on θ_2 (verbal ability)	0.19**	0.00	0.34**	1.48**	1.00**	0.18**	0.00	0.35**	1.79**	1.00**	0.19**	0.00	0.37**	2.24**	1.00**
σ_ϵ^2	11.3**	10.6**	25.5**	20.2**	50.9**	10.2**	12.5**	25.3**	19.5**	55.0**	9.10**	12.3**	24.0**	16.6	53.8**

Column titles: GPA – high school GPA, M – math score, S – science score, SS – social sciences score, T – Turkish score. Significance levels: * – 5%, ** – 1%. Standard errors are obtained by bootstrapping (15,000 bootstrap samples used).

Table 8. Estimates of the GPA and the Score Equations.

Income, I	Low	Medium	High
Expected learning shock, $\mu_{\lambda t}(I)$			
$t = 2$	7.05**	9.34**	6.2**
$t = 3$	2.59**	1.66*	-1.77
$t = 4$	-1.30	3.00	1.06
Variance of the learning shock, $\sigma^2_{\lambda t}(I)$			
$t = 2$	189**	216**	275**
$t = 3$	98**	134**	182**
$t = 4$	17	0	0
Costs of retaking, $\psi_t(I)$			
$t = 1$	0.0282**	0.0420**	0.0157
$t = 2$	0.0023	-0.0049	-0.0261
$t = 3$	-0.0182	0.0030	-0.0086
$t = 4$	0.0015	0.0067	-0.0109

Significance levels: * – 5%, ** – 1%. Standard errors are obtained by bootstrapping (15,000 bootstrap samples used).

Table 9. Estimates of the Structural Parameters: Learning Shocks and Retaking Costs

the differences are only marginally significant. If anything, the gambling motive works against higher retaking rates observed among the low-income students.

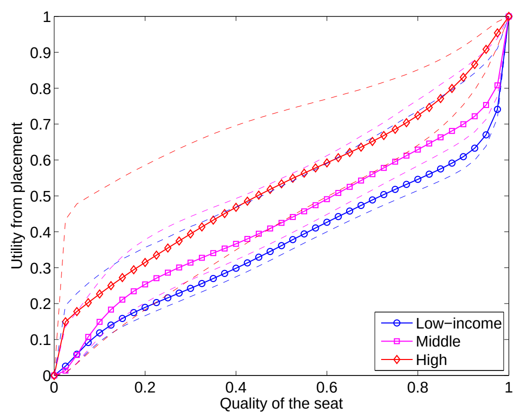
5.3 Payoffs from Placement and Costs of Retaking

Finally, in step 3, we estimate the utility function, $U(r, I)$, and the retaking costs, $\psi_t(I)$. Recall that the best placement is normalized to unity and the worst to zero. Retaking costs, which may be pecuniary or non-pecuniary, are larger for the first retaking attempt than for later ones, where they are not statistically different from zero (see Table 9).

Retaking costs in the first retaking attempt correspond to roughly 7.5 percent of the utility of an average placement. Surprisingly, retaking costs are not significantly different across income groups. For example, these costs could be higher on the first retake due to a stigma attached to retaking in itself.

The non-parametric estimates of the utility functions $U(r, I)$ are depicted in Figure 4. Note that, at the top, the marginal benefit of a higher score is low for the rich, but high for the less well-off. This makes the poor who score at the top more risk loving than the rich with similar scores. This could be explained by the rich caring less about getting into the best schools. Since their future success depends less on their exam performance due to better outside options (e.g., starting a business or joining the family firm) when compared to the poor, students who are better off benefit little from moving from an already highly-ranked placement to a marginally better one.

Figure 4. Estimated Utility by Income Group



6 Counterfactual Experiments

We conduct a number of counterfactual experiments below, all aimed at reducing retaking. Our objective is to predict the consequences of various reforms so as to understand the trade-offs involved and the distributional effects that each of them may entail.[44]

Table 10 compares the simulation results from the current scenario with unlimited retaking to a scenario that forces all students to take the exam in the last year of high school and forbids retaking (1 attempt) and a scenario where a maximum of two attempts is allowed (2 attempts). We also look at the consequences of penalizing retakers by reducing their scores by 5 percent ("5% penalty"). Finally, we experiment with doubling the weight on GPA ("x2 GPA") in the placement process.

	Income	Current	1 attempt	2 attempts	5% penalty	x2 GPA
Payoffs	low	0.27	0.31	0.29	0.30	0.28
	medium	0.38	0.42	0.40	0.41	0.39
	high	0.55	0.59	0.57	0.59	0.56
# of attempts	low	2.28	1.00	1.40	1.52	1.97
	medium	1.94	1.00	1.35	1.38	1.69
	high	1.88	1.00	1.30	1.37	1.65
Ability before attempt 1						
% underplaced	low	32.82	15.55	23.18	21.45	26.13
	middle	35.10	16.92	25.61	22.16	28.63
	high	41.94	16.74	29.83	21.75	32.71
Ability at placement						
% underplaced	low	8.44	15.55	12.65	13.65	6.03
	middle	10.24	16.92	12.81	16.34	8.03
	high	12.61	16.74	13.19	15.78	9.92

Policies: current – unlimited retaking, 1 attempt max, 2 attempts max, 5% penalty after attempt 1, the weight on GPA is doubled. We use endogenous admission cutoffs in all counterfactuals.

Table 10. Policy Experiments.

6.1 Effect on Overall Payoffs

Retaking is costly both in terms of direct costs incurred by students, as well as in terms of their effect in equilibrium. Recall that the private benefit from retaking exceeds the social benefit so that retaking is excessive. As more people retake, cutoff scores for admission are bound to be higher, both because of the larger numbers involved and because of learning between attempts. Payoffs are defined as the expected utility of placement less costs of retaking. Table 10 shows that under the

[44] In Appendix A.3, we use the extended model that allows for endogenous schooling choices before the first attempt and consider only a ban on retaking. Note, this is the only counterfactual we can consider that is not subject to the Lucas Critique. Our extended model delivers very similar results as the baseline version.

current system payoffs are increasing in income. This comes from higher-income students tending to have higher scores and therefore better placements. In addition, they tend to retake less often, which reduces their costs. As we look across policies, it becomes apparent that preventing retaking results in higher welfare than any other policy for each income group. The reason for this is the negative externality generated by retakers, which results in excessive retaking.

6.2 Effects on Underplacement

Discouraging retaking may result in students being mismatched with schools in terms of their ability. In settings where there are social benefits from matching better students with better schools, something we do not explicitly model, discouraging retaking may have significant costs. Nevertheless, we can get some insight into this by looking at the consequences of these policies on the extent of underplacement. We consider a person to be underplaced if the ranking of the placement he obtains is more than 5 percent below the person's rank in terms of the basis used—ability in attempt 1, or ability at the time of placement. Defining underplacement in terms of ability at placement makes sense as this is the most recent measure of ability. On the other hand, using this definition makes the ability depend on the counterfactual. One might want to therefore define underplacement in terms of a fixed basis using initial ability. Initial ability and ability at placement are proxied by the noise-free score in the first attempt and at the time of placement, respectively.

Three patterns emerge from Table 10. First, the fraction underplaced is higher with retaking when we use initial ability to define underplacement. Second, the opposite happens when underplacement is defined in terms of ability at placement. Third, the rate of underplacement rises with income in the current system independent of which definition we use.[45]

What explains these patterns? Consider underplacement according to initial ability. Students are placed according to their rank (in terms of score), so that anything that reduces the informativeness of the score upon placement, also increases the extent of underplacement. With random learning shocks, this is exactly what happens. When more retaking is allowed, the extent of underplacement in terms of initial ability tends to increase. This explains the first pattern. This force does not operate when underplacement is defined in terms of ability upon placement.

When underplacement is defined in terms of ability at placement, the insurance aspect of retaking drives the pattern observed. Under the current system, bad transitory shocks merely result in retaking, so that underplacement rates are low. Restricting retaking pushes agents to accept their placement, even if it is bad, which raises the extent of underplacement. This explains the second pattern.

Finally, if the net costs of retaking vary by income, the option to retake will be differentially exercised: high net cost groups will retake only if the negative shock to score is very large and so

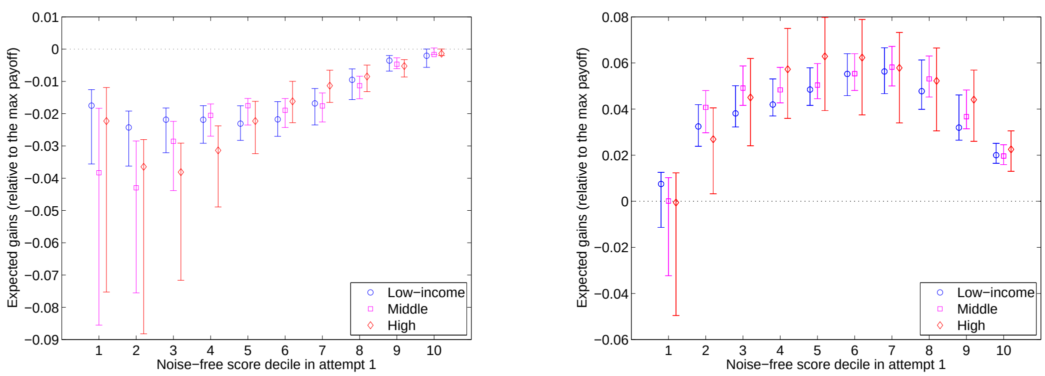
[45] Changing the definition of underplacement to 20 percent or 1 percent results in similar patterns.

will tend to be underplaced more often. This seems to be what lies behind underplacement rates increasing with income, since those with higher incomes retake less.

6.3 Welfare Gains and Ability

We have seen so far that limiting, or eliminating retaking improved expected welfare for each of the three income groups. Of course, there is considerable heterogeneity within each income group. Next, we look at how these welfare gains vary by ability as captured by their initial score decile.

Figure 5. Gains From Banning Retaking



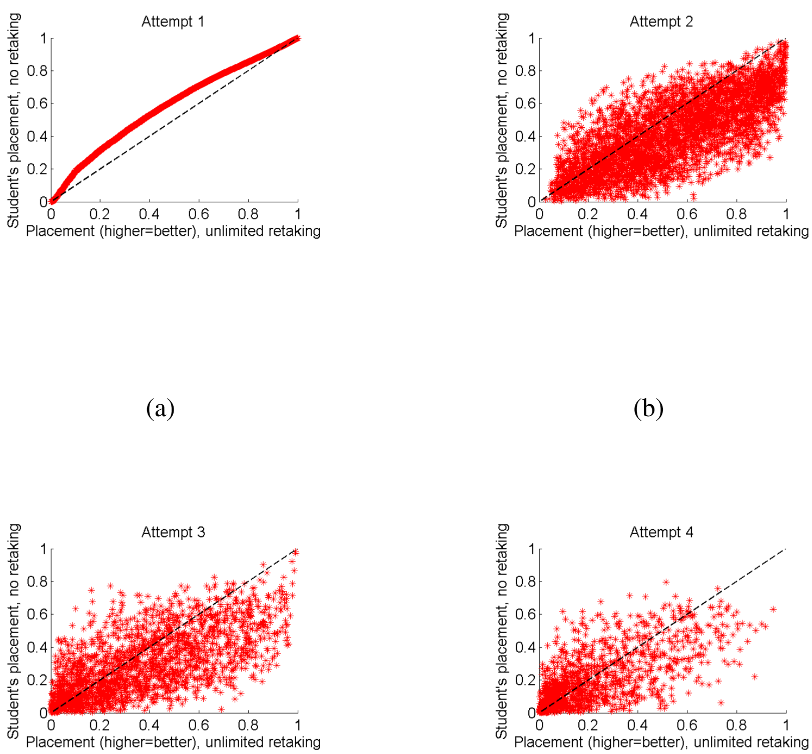
(a) Admission cutoffs are kept fixed

(b) Admission cutoffs adjust endogenously

A naive agent would assume that the admission cutoffs are fixed. Under this assumption, we look at the expected payoff gains/losses from preventing retaking. As shown in panel (a) of Figure 5, students in low initial score deciles lose more. This should be expected, as retaking tends to decrease in score among first-time takers in our data. Thus, low initial score students lose the most when retaking is banned. However, as pointed out earlier, the fallacy of composition is at work. For each student, it is better to be allowed to retake than not, given the cutoff scores. However, if all students are prevented from retaking, then the cutoff scores fall. This general equilibrium effect reverses the welfare effects of banning retaking, as depicted in panel (b) of the same figure. Again, this makes sense as there is excessive retaking due to the externality identified earlier.

The general equilibrium effects are illustrated in Figure 6. Each student's placement under the no-retaking rule is plotted against his placement in the current system, for attempts 1–4. The upper-left panel of the figure depicts students who are placed in the first attempt under the current system with what their placement would have been in equilibrium had retaking been banned. The curve is above the 45 degree line showing that the cutoffs fall (quality of placements rise for the same score) when retaking is banned. Moreover, the fall in the cutoffs is greatest for those with mid-range placements. As the number of attempts rises, the mass moves towards the origin as low-ability students retake more often and are placed in low-quality seats under either policy.

Figure 6. Placement with and without Retaking



(a) (b)

(c) (d)

In sum, the partial equilibrium consequences of preventing retaking are to reduce welfare for everyone, and more so for those with low scores. Nevertheless, welfare gains are observed in the general equilibrium, and more so for those in the middle score deciles, as is evident from the top-left corner of Figure 6, where cutoffs rise only in the middle. The second effect dominates, resulting in inverse U-shape gains in Figure 5b. Though most agents gain ex post, about 20 percent of them lose.[46] Some idea of this can be gleaned from Figure 6 as a significant number of students are below the 45 degree line, which means their placement is worse with no retaking. However, the figure does not capture welfare changes fully, as the lower expenditures on retaking are not accounted for. Taking these costs into account will reduce the number of losers ex post.

Redistributional effects across income groups seem to arise mostly through differences in initial performance. Figure 5b shows that differences in gains across income groups are not significant after controlling for the initial score decile. Although income groups do have different learning effects upon retaking as well as different retaking costs, these effects seem to wash out.

7 Conclusion

In this paper, we have documented that, at least for the setting we examine, limiting retaking, though seemingly harmful to individuals, is in their interest in equilibrium. This stark contrast between individual incentives and aggregate ones suggests that reform in this arena may be difficult to implement. Individuals will naturally resist attempts to reduce the options open to them as general equilibrium effects tend to be opaque. By quantifying the full effects of a reform in a general equilibrium setting, we can identify win-win policies such as limiting retaking that will probably face opposition ex ante.

In our analysis, we have, of course, made some simplifying assumptions. First, we assume that preferences are vertical. Our focus is on retaking, not preferences, so that simplifying the latter to zoom in on the former is natural.[47] We model utility as increasing in the score/rank of an agent. This would be true even if preferences were horizontal as a higher score makes more options available to a student.

Second, we do not account for active learning. In our model learning is a draw from a distribution that an agent takes as given. By choosing to retake, the agent can choose to draw from this distribution but cannot choose the distribution he draws from by, say, expending effort. Thus, we are not able to distinguish between fixed and variable (effort) costs of retaking in our estimates. We are, however, able to incorporate investment into schooling and tutoring prior to the

[46] By ex post we mean that we keep the shocks faced by agents constant across policy scenarios.

[47] Purely vertical preferences may be problematic if we were studying other questions. For example, had we been looking at the effects of expanding certain schools it would be important to know substitution patterns in demand, and imposing vertical preferences would constrain these patterns significantly. However, detailed information about substitution patterns seems less vital in modeling retaking.

first attempt. We find that ex ante welfare rises with a ban for most agents, though the size of the welfare gain is roughly halved.

Third, we focus on steady state outcomes. The welfare consequences out of steady state are likely to be different. In particular, if retaking is banned and the policy is unexpected, then those who planned to retake would suffer considerably. Thus, implementation would have to be gradual and exempt previous cohorts, which would then reduce or eliminate welfare gains for them. The precise timetable involved would be critical in determining out of steady state welfare gains/losses. A better understanding of these trade-offs is a topic for future work.

References

AGUIRREGABIRIA, V., AND P. MIRA (2010): "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*, 156(1), 38–67.

AKERLOF, G. (1976): "The Economics of Caste and of the Rat Race and Other Woeful Tales," *The Quarterly Journal of Economics*, 90(4), 599–617.

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

BALINSKI, M., AND T. SÖNMEZ (1999): "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory*, 84(1), 73–94.

BONHOMME, S., AND J.-M. ROBIN (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *The Review of Economic Studies*, 77(2), 491–533.

CANER, A., AND C. OKTEN (2010): "Risk and Career Choice: Evidence from Turkey," *Economics of Education Review*, 29(6), 1060–1075.

——— (2013): "Higher Education in Turkey: Subsidizing the Rich or the Poor?," *Economics of Education Review*, 35, 75–92.

DEMING, D., AND S. DYNARSKI (2008): "The Lengthening of Childhood," *The Journal of Economic Perspectives*, 22(3), 71–92.

FAIN, J. R. (2009): "Affirmative Action Can Increase Effort," *Journal of Labor Research*, 30(2), 168–175.

FREYBERGER, J. (2012): "Nonparametric Panel Data Models with Interactive Fixed Effects," University of Wisconsin.

FRISANCHO, V., K. KRISHNA, S. LYCHAGIN, AND C. YAVAS (2016): "Better Luck next Time: Learning through Retaking," *Journal of Economic Behavior & Organization*, 125, 120–135.

FU, Q. (2006): "A Theory of Affirmative Action in College Admissions," *Economic Inquiry*, 44(3), 420–428.

HATAKENAKA, S. (2006): "Higher Education in Turkey for 21st Century: Size and Composition," Discussion paper, World Bank.

HOTZ, V. J., AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60(3), 497–529.

MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816.

MANKIW, N. G., AND M. D. WHINSTON (1986): "Free Entry and Social Inefficiency," *The RAND Journal of Economics*, 17(1), 48–58.

POLYANIN, A. D., AND A. V. MANZHIROV (2008): *Handbook of Integral Equations: Second Edition*. CRC Press.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

SAYGIN, P. O. (2016): "Gender Differences in Preferences for Taking Risk in College Applications," *Economics of Education Review*, 52, 120–133.

STIGLER, G. J. (1961): "The Economics of Information," *Journal of Political Economy*, 69(3), 213–225.

TANSEL, A., AND F. BIRCAN (2005): "Effect of Private Tutoring on University Entrance Examination Performance in Turkey," IZA Discussion Paper 1609.

TÖRNKVIST, B., AND W. HENRIKSSON (2004): "Repeated Test Taking: Differences between Social Groups," Discussion Paper 47, Umea University.

TRAIN, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press.

VIGDOR, J. L., AND C. T. CLOTFELTER (2003): "Retaking the SAT," *The Journal of Human Resources*, 38(1), 1–33.

A Sensitivity Checks

A.1 Normalizing the High School GPA

The placement score defined in equation 4 depends on the normalized high school GPA, g_i. Unfortunately, we observe the raw and the standardized GPA, but not g_i. While we know the official formula for g_i, it requires other missing information. In this section, we describe the simplified normalization we use and then show that this seems to be close to what is actually used in placement and that our simplification does not qualitatively affect the estimates.

A.1.1 The Official Formula

First, consider the official formula for normalized GPA. Let i denote a student who graduates from school k's science track. The exam authority computes this student's GPA in two steps. First, his raw GPA is rescaled so that the average GPA within the school equals 50 and the standard deviation equals 10. Values above 80 (below 30) are top-coded (bottom-coded, respectively). The resulting standardized GPA is denoted as GPA_{ik} and is documented in our data for every student. In the second step, the exam authority computes the normalized GPA, g_{ik}, using the following formula:

$$\begin{aligned} g_{ik} &= gmin_k + (80 - A_k)\frac{GPA_{ik} - GPAMIN_k}{GPAMAX_k - GPAMIN_k} \\ gmin_k &= 8 + (GPAMIN_k - 8)\,\frac{ASCORE_k}{80} \end{aligned} \tag{13}$$

Students who get the highest GPA in school, $GPAMAX_k$, obtain 80 points (the maximum). Students with the lowest GPA, $GPAMIN_k$, obtain $gmin_k$. The latter puts a lower bound on the normalized GPA, which increases in $ASCORE_k$, school k's average of $\sum_j w_j s_{ij}$, the score accepted by college programs in natural sciences. Thus, students whose classmates perform better at the entrance exam get a larger boost to their normalized GPA, which accounts for higher quality of their schools.

While the formula in (13) links g_i to the standardized GPA, which we observe, it also requires $GPAMIN_k$ and $GPAMAX_k$, neither of which are present in the data.[48] We follow a two-pronged approach to overcome this issue. In our main specification, we use a simplified normalization that approximates the official formula without relying on $GPAMIN_k$ or $GPAMAX_k$. In Section A.1.3, we estimate the official GPA for students who graduate from relatively larger and better schools. For this subsample, we demonstrate that the estimates are not sensitive to whether we use the simplified formula or the official one.

[48] We can compute the minimum and the maximum in our sample of students, but not in the population.

A.1.2 Our Simplified Formula

Raw GPAs ignore potential quality heterogeneity and grade inflation across high schools. However, obtaining 10/10 at a very selective school is not the same as obtaining 10/10 at a very bad school. The official formula in equation (13) corrects for this by raising the GPAs of students from better-performing schools.

Since we cannot use the official formula, we correct for grade inflation in a different way, following Frisancho, Krishna, Lychagin, and Yavas (2016). For each school k, we define the adjustment factor, A_k:

$$A_k = \frac{ARAWGPA_{sc,k}}{ASCORE_{sc,k}} \frac{ASCORE_{sc}}{ARAWGPA_{sc}} \tag{14}$$

where $ARAWGPA_{sc,k}$ and $ASCORE_{sc,k}$ are the average raw GPA and the average weighted exam score, $\sum_j w_j s_{ij}$, in the sample of students who originate from the science track in school k.[49] The weights on the subject scores, w_j, are the same as used by the exam authority to calculate Y-ÖSS-SAY scores. $ARAWGPA_{sc}$ and $ASCORE_{sc}$ are the average raw GPA and weighted score across all science track students in the data.

The first ratio in (14) should go up if the school is inflating grades relative to its true quality. For example, if the average GPA in school j is about 8/10 but the average exam score for its students is only 5/10, school j is worse than the raw GPAs of its students suggest. The second ratio in (14) is just a constant that takes the adjustment factor to a scale relative to everyone in the science track.

We define the normalized GPA for student i in school k as

$$g_{ik} = \frac{RAWGPA_{ik}}{A_k} \frac{100}{\max_{i,k} \left(RAWGPA_{ik}/A_k \right)}$$

where the maximum in the denominator is found using the observed sample of science students, not the population. Note that if school k tends to inflate grades relative to true performance, the raw GPA of all the students in k will be penalized through a higher A_k.

A.1.3 Official vs Simplified Normalization: Sensitivity Checks

Despite the fact that we do not observe all components of the official formula in (13), we can find tight bounds on the official GPA for approximately 60 percent of students in our final sample. We do so using preference lists submitted by the students, their placements and the cutoff scores for each program.

[49] Note that $ASCORE_{sc,k}$ is different from $ASCORE_k$ used in the official formula (13). The former is the average exam score of science track students from the sample, while the latter includes the population of students from all tracks in the high school k. For schools that are represented by less than 5 students in our sample, we used the average score and GPA of students who attended the same school type (i.e., public, private, Anatolian/science school).

Our approach is best illustrated using an example. Suppose that we have a science student whose weighted score $\sum_j w_j s_{ij}$ is 150 points. The student applies to three programs compatible with the science track: A, B and C, in the order of preference. We observe that he was offered admission at B. We also observe the realized admission cutoffs: $S_A = 195$, $S_B = 175$, $S_C = 170$. Since A, B and C are compatible with the student's high school track, the weight on HS GPA in equation (4) equals 0.5. Therefore, the student's placement score is $s_i = 150 + 0.5g_i$, where g_i is unknown. While one cannot find the precise value of g_i in this example, the data above provides bounds on g_i. First, since the student was not accepted to A, his placement score is below A's cutoff: $150 + 0.5g_i < 195$. Second, the student's score is above B's cutoff: $150 + 0.5g_i \geq 175$. The cutoff at C is irrelevant, since the student is accepted to B. Together, this implies that $50 \leq g_i < 90$. Another set of constraints comes from equation (13): normalized GPA cannot be greater than 80 or less than 30 points. Taking this into account, the non-parametric bounds on g are as follows: $50 \leq g \leq 80$.

The non-parametric bounds are fairly tight only for a tiny fraction of students. For this reason, we impose the parametric structure of equation (13). For each school k, we run a grid search in the space of all admissible pairs $(GPAMIN_k, GPAMAX_k)$, use (13) to generate candidate g_i's for all students from school k, and save the pairs that violate the smallest number of non-parametric restrictions. Then, every such pair is used to generate g_i, resulting in an individual set of g_i's for each person who graduated from school k. The minimum and the maximum values in this set provide parametric bounds on g_i. If these bounds are at most 2 points apart, the midpoint between them approximates the true g with an error of at most one point. This level of precision is attainable for approximately 60 percent of students in our sample. In what follows, we refer to these students as the "official GPA sample," while the data used in the main body are labeled as the "full sample."

Table 11 shows that the official GPA sample of students is systematically different from the full sample in terms of the high school type. The bounds on the student's GPA are tight only if large numbers of other students from the same school submit long preference lists and are present in our data. Since private schools tend to have smaller classes than other school types, we cannot back out GPAs for many private school graduates. Table 11 also demonstrates that students in the official GPA sample tend to score slightly better than those in the full sample, but their performance in high school seems to be almost identical. GPAs computed using the simplified and the official formulas have very similar means and standard deviations. They are also tightly related to each other: the correlation coefficient of the two is above 0.9.

In order to check if our structural estimates are sensitive to the definition of GPA, we run our estimation code three times. First, we use the full sample and our simplified definition of GPA. Second, we switch to the official GPA sample, but keep using the simplified GPA. This allows

Sample	Full sample		Students with official GPA	
	mean	st.dev.	mean	st.dev.
High school type				
Anatolian/science	0.22		0.30	
Public	0.56		0.55	
Private	0.19		0.13	
Shares of students in each income group				
Low	0.32		0.30	
Medium	0.41		0.41	
High	0.27		0.30	
Weighted exam score	127.66	22.72	130.00	23.53
Normalized GPA				
Simplified formula	56.53	11.97	56.78	12.72
Official formula			59.42	10.20
No. of obs.	31,554		18,873	

Full sample — data used to obtain the estimates in the main body. Official GPA — subset of the full sample for which we can obtain official GPA with 1-point precision.

Table 11. All students vs students with precise data on normalized GPA.

us to control for changes in the estimates stemming from the compositional differences between the two samples. Third, we use the official GPA sample together with the official GPA definition. The estimates are reported in Table 12. The consequences of switching from the simplified formula to the official one are found by comparing specifications (2) and (3). The estimates in the respective columns tend to be very close in all three income groups. The few parameters that vary substantially between the specifications are identified by relatively small subsamples of students (for instance, $E[\lambda_4]$ and $Var[\lambda_4]$ in the low-income group are pinned down by data on less than 500 low-income fourth-time takers).

In our final robustness check, we compare the welfare implications of banning retaking for the three income groups. We numerically simulate the current and the no-retaking policy in the three specifications defined above. The results are reported in Table 13. While the levels of predicted utility vary between the specifications, the estimated gains from the ban on retaking are always positive and have the same magnitude.

Overall, the above estimates suggest that our results are not sensitive to replacing the official GPA with its simplified version. The official GPA can be reliably estimated only for a smaller subsample of the original dataset. This subsample is also systematically different from the rest of the student population. For these reasons, we use the simplified formula in our main analysis.

Income group	Low			Medium			High		
Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Expected learning shock, $E[\lambda_t]$									
attempt 2	7.05	7.17	7.08	9.34	8.51	7.99	6.20	6.78	6.38
attempt 3	2.59	3.12	1.50	1.66	1.55	-1.79	-1.77	-2.43	-2.55
attempt 4	-1.30	-0.68	-4.29	3.00	3.08	1.10	1.06	1.01	1.73
Variance of the learning shock, $Var[\lambda_t]$									
attempt 2	189	144	192	216	237	242	275	265	260
attempt 3	98.2	149	168	134	110	149	182	148	132
attempt 4	16.7	20.3	0	0	1.54	0	0	0	0
Cost of retaking, ψ_t									
attempt 1	0.028	0.030	0.027	0.042	0.034	0.045	0.016	-0.018	-0.011
attempt 2	0.002	0.003	-0.014	-0.005	-0.014	-0.025	-0.026	-0.058	-0.055
attempt 3	-0.018	-0.023	-0.050	0.003	0.002	-0.006	-0.009	-0.034	-0.025
attempt 4	0.001	0.002	0.000	0.007	0.007	0.005	-0.011	-0.032	-0.028

Specifications: (1) Estimates from the main body of the paper — full sample, GPA is computed using the simplified formula. (2) Subsample of students with precise estimates of the official GPA, but the GPA is still computed using the simplified formula. (3) Subsample with precise estimates of the official GPA, GPA is computed using the official formula.

Table 12. Robustness of structural parameter estimates to the choice of the GPA formula

A.2 Excluding Employed Applicants

In our main empirical exercise we treat employed applicants as regular exam takers. However, compared to those who prepare for the exam full time, employed retakers may be more willing to forego placement and have less time to prepare for the exam. They may also be targeting very different programs compared to the rest of the population. For instance, distance programs, which are not very desirable by the majority, may be preferred by the working applicants, as it is easier to keep a job and earn a degree in a distance program than in a full-time one.

Statistics on exam scores and placements confirm our intuition. Table 14 shows that working students predominantly go to programs with non-binding quotas, unlike students who are not employed at the time of taking the exam. Table 15 demonstrates that working students also tend to get substantially lower scores. After controlling for demographics and performance at high school, the gap between students who do and do not work reaches almost 13 points, which is roughly the same size or higher than the estimated learning gains in Table 9. This is in line with what one would expect if students working full-time had no opportunities to improve or even perform at the same level as in their first attempt.

For the above reasons, one may argue that working applicants should be treated in the same way as we treat enrolled college students. That is, the choice to become employed should be

Specification	(1)		(2)		(3)	
Policy	Current	1 attempt	Current	1 attempt	Current	1 attempt
Expected ex-ante utility, by income						
Low	0.27	0.31	0.33	0.36	0.35	0.37
Middle	0.38	0.42	0.42	0.45	0.36	0.40
High	0.55	0.59	0.71	0.74	0.68	0.70

Specifications: (1) Estimates from the main body of the paper — full sample, GPA is computed using the simplified formula. (2) Subsample of students with precise estimates of the official GPA, but the GPA is still computed using the simplified formula. (3) Subsample with precise estimates of the official GPA, GPA is computed using the official formula. Policies: current – unlimited retaking, 1 attempt max

Table 13. Robustness of counterfactual estimates to the choice of the GPA formula

viewed as an irreversible decision that removes the student from the population of interest as he is extremely unlikely to compete for access to competitive programs with the rest of the science track.

We perform a series of sensitivity checks by excluding working students from the main sample. In our first exercise, we exclude all exam takers who are employed full time, re-run our estimator and the counterfactual simulations. According to Table 14 and 15, these students differ the most from the rest of the population in terms of preferences and performance. Then, we take an even more conservative approach and exclude all working students.

The estimates of learning shocks and retaking costs from both exercises are reported in Table 16, along with the baseline estimates from the main body of the paper. Consistent with the preliminary evidence, expected learning shocks tend to go up as we exclude working students. The general pattern remains the same: the largest improvements in scores come in the early attempts, and students from middle income families improve the most. Estimated costs of retaking go up. This happens because now we treat work as an irreversible decision to stop competing in the same market for seats as the rest of the population. Thus, we count employed students as non-retakers even when they do retake: they do so to access programs with quota-free admission. To produce lower retaking rates the model has to have higher retaking costs.

As Table 17 demonstrates, the main outcome of the policy simulations remains the same as before. A ban on retaking raises ex ante expected utility for exam takers in all three income groups.

A.3 Incorporating Pre-Test Investment

If retaking is restricted, the students may turn to other costly ways of competing in their exam scores. For instance, students who used to go to public schools under the unlimited retaking policy may enroll in fee-paying high schools and pay for private tutoring after retaking is banned. To

Percentage of students placed in 2002	Program of placement: non-binding quota	binding quota	Total
Not working	15.2	73.4	88.6
Working, temporarily	1.7	1.5	3.2
Working, part-time	0.6	0.8	1.4
Working, full-time	6.1	0.8	6.8
Total	23.5	76.5	100

Table 14. Placements of Science Track Students by Their Employment Status.

Dependent variable: ÖSS-SAY score	Attempts 1	2	3	4	5+
Working, full-time	-4.74*	-12.95**	-12.30**	-7.33**	-4.95**
	(2.16)	(0.88)	(0.63)	(0.69)	(0.82)
Working, part-time	-1.35	-4.65**	-5.21**	-2.70	-1.30
	(0.97)	(1.30)	(1.42)	(1.68)	(1.58)
Working, temporarily	-2.23**	-6.32**	-8.63**	-6.03**	-4.30**
	(0.84)	(0.87)	(0.90)	(0.92)	(0.92)

Notes: Each column reports coefficients from a regression that has ÖSS-SAY score as a dependent variable. The coefficients are estimated via OLS independently for each attempt and pooling all income groups. The unreported controls include all variables from Table 8, income dummies, high school GPA, its quadratic and cubic term. Robust standard errors are reported in the parentheses. Students who are not working at the time of taking the exam form the baseline category.

Table 15. Student Employment Status and Exam Performance.

explore this possibility, we augment our model with a stage that describes the choice of high school type and additional tutoring.

We modify the timing of the model in Subsection 3.3 by adding a time period zero, taking place between middle and high school. In this period, every student from our population[50] makes a costly investment in schooling. The student has to choose among three broad categories of high schools: public, private and Anatolian/science. Student i's choice is denoted as $hs_i \in HS = \{PUB, PRIV, ANAT\_SC\}$. The student also makes a decision whether to get private tutoring for the entrance exam; $pt_i \in \{0,1\}$ denotes i's choice. In total, each student faces six options: public school with no tutoring, public school with tutoring, private school with no tutoring, and so on.

Before making the choice (hs_i, pt_i), student i observes X_{0i}, a vector of his own socioeconomic characteristics that contains the same components as X_i excluding high school type and

[50] Note that we do not model behavior of students who are not college-bound, who end up in schools incompatible with the science track (e.g., vocational or religious schools) or who end up in other tracks. As before, our population of interest includes only college-bound students from the science track.

Income group	Low			Medium			High		
Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Expected learning shock, $E[\lambda_t]$									
attempt 2	7.05	7.84	8.52	9.34	10.2	10.7	6.20	7.55	7.86
attempt 3	2.59	4.05	4.37	1.66	3.33	4.39	-1.77	2.78	4.13
attempt 4	-1.30	-0.99	-1.98	3.00	1.91	3.00	1.06	0.35	-2.69
Variance of the learning shock, $Var[\lambda_t]$									
attempt 2	189	183	209	216	186	151	275	232	208
attempt 3	98	147	138	134	154	194	182	226	222
attempt 4	17	29	48	0	8	48	0	0	20
Cost of retaking, ψ_t									
attempt 1	0.028	0.032	0.038	0.042	0.052	0.055	0.016	0.033	0.043
attempt 2	0.002	0.015	0.016	-0.005	0.012	0.022	-0.026	0.019	0.035
attempt 3	-0.018	-0.017	-0.018	0.003	0.001	0.010	-0.009	0.000	-0.002
attempt 4	0.001	0.001	0.001	0.007	0.008	0.008	-0.011	0.000	0.006

Specifications: (1) Estimates from the main body of the paper — all working exam takers are included. (2) Exam takers with full-time jobs are excluded. (3) Exam takers with full-time, part-time or temporary jobs are excluded.

Table 16. Sensitivity of Structural Estimates to the Exclusion of Employed Exam Takers

Specification	(1)		(2)		(3)	
Policy	Current	1 attempt	Current	1 attempt	Current	1 attempt
Expected ex-ante utility, by income						
Low	0.27	0.31	0.29	0.33	0.29	0.33
Middle	0.38	0.42	0.36	0.41	0.37	0.42
High	0.55	0.59	0.51	0.57	0.49	0.55

Specifications: (1) Estimates from the main body of the paper — all working exam takers are included. (2) Exam takers with full-time jobs are excluded. (3) Exam takers with full-time, part-time or temporary jobs are excluded.

Table 17. Sensitivity of Policy Simulations to the Exclusion of Employed Exam Takers

variables related to tutoring. The student also knows g_{0i}, his middle-school GPA that takes one of three values (A, B or C).

A.3.1 Costs of Pre-Test Investment

Schooling choices are associated with costs, which capture the fees and effort of keeping up with curriculum. Choosing high school category hs and private tutoring pt entails a cost of

$$c_i(hs, pt) = C(I_i)\left[\sum_{g_0\in\{A,B,C\}}\sum_{I=1}^{3}\gamma_{hs,pt,g_0,I}\mathbf{I}[g_{0i}=g_0, I_i=I] + u_{i,hs} + v_{i,pt} + w_{i,hs,pt}\right],$$

$$hs \in \{PUB, PRIV, ANAT\_SC\},\ pt \in \{0,1\}, \quad (15)$$

where $\mathbf{I}[g_{0i} = g_0, I_i = I]$ denotes an indicator that equals one if and only if student i's middle school GPA equals g_0 and if i belongs to the income group I. The coefficients γ vary across choices (hs, pt).

The first component of the cost in (15) captures dependence on individual observables. For instance, expensive private schools may be prohibitively costly for students from lower-income families. This would be reflected in a higher value of $\gamma_{PRIV,pt,g_0,I}$ for $I = 1$ (i.e., low income) compared to that for $I = 3$ (high income). Middle-school GPA g_0 accounts for the level of preparation; students who are lagging relative to their peers may have harder time getting a seat and graduating from a more demanding school.

Schooling costs also depend on unobservables, which are captured in (15) by the three student-specific shocks: $u_{i,hs}$, $v_{i,pt}$ and $w_{i,hs,pt}$. The first two, u and v, reflect costs specific to private tutoring and formal schooling, respectively. Public school with no tutoring is set as the baseline option so that $\gamma_{PUB,0,g_0,I} = u_{i,PUB} = v_{i,0} = 0$. The remaining $[u_{i,PRIV}, u_{i,ANAT\_SC}, v_{i,1}]$ are jointly normal with a zero mean and a covariance matrix that is allowed to be different for each income group, $\Sigma_c(I_i)$. Idiosyncratic shocks $w_{i,hs,pt}$ are independently drawn from the standard Gumbel distribution for each choice and student.

Finally, the costs are multiplied by $C(I_i)$, a factor that converts cost units into placement utility units.[51] This factor is allowed to vary across income groups: higher-income students may have lower marginal valuation of money spent on tutoring relative to improvements in placement outcomes that this money can buy.

A.3.2 Payoffs

Choosing a more expensive school type can affect performance in the college exam as each schooling option influences the student's expected noise-free score in the first attempt, $\bar{s}_{i1}$. Obtaining a

[51] One unit of placement utility is equal to the payoff of getting the best seat. Cost units are chosen so as to normalize the variance of $w_{i,hs,pt}$ to $\pi^2/6$, the variance of the standard Gumbel distribution.

higher score, in turn, raises student's expected utility, $E[V_1(\bar{s}_{i1}, \varepsilon_{i1}, I_i)]$. We map this correspondence between choices and payoffs more formally below.

According to equations (2) and (4), student i's score in attempt 1 is[52]

$$s_{i1} = X_i' \left(\frac{\beta_g(I_i)}{2} + \beta(I_i) \right) + \theta_i' \left(\frac{\alpha_g(I_i)}{2} + \alpha(I_i) \right) + \frac{\varepsilon_{ig}}{2} + \varepsilon_{i1}$$

Note that the coefficients β in this equation do not have causal interpretation, as students with higher innate ability θ_i may self-select into better schools and receiving private tutoring. At the time of choosing (hs, pt), the student does not observe $[\theta_i' (\alpha_g(I_i)/2 + \alpha(I_i)) + \varepsilon_{ig}/2]$; he only knows his middle-school GPA, g_{0i}, as a noisy signal of the former term. Part of the vector X_i, X_{0i}, is observed and treated by the student as fixed; the remaining part is related to the choice of (hs, pt) and is thus endogenous.

Let $s_{i1}(hs, pt)$ be the score that student i gets in the first attempt if he chooses (hs, pt). We assume that student's beliefs about the future score in the first attempt are described by the following equation:

$$s_{i1}(hs, pt) = \sum_{g_0 \in \{A,B,C\}} \sum_{I=1}^{3} \mathbf{I}[g_{0i} = g_0, I_i = I] \left(\rho_{hs,pt,g_0,I} + X_{0i}' \chi_{g_0,I} \right) + \varepsilon_{i0} + \varepsilon_{i1}$$

That is, i's Y-ÖSS-SAY score can be partly predicted using the student's family background, X_{0i} and I_i, and his middle-school GPA, g_{0i}. The shock ε_{i0} captures the part of s_{i1} that is not known before attending high school, but is known just before the first attempt at the exam; ε_{i1} is drawn at the exam as mentioned in the main body of the paper and is independent of any variables generated previously, including X_{0i}, g_{0i} and ε_{i0}. The standard deviation of ε_{i0}, $\sigma_{\varepsilon 0}$, is allowed to depend on student's income group and middle-school GPA, I_i and g_{0i}.[53]

Assumption 6. *When student i chooses (hs, pt), he is completely uncertain about the value of ε_{i0}.*

To put it differently, this assumption implies that there is no unobserved ability on the right hand side of equation (**??**) after controlling for student's middle-school GPA and socioeconomic characteristics. As a result, there is no selection on ε_{i0} when students decide on high school type and tutoring. To predict his future score conditional on choosing (hs, pt), student i looks at the scores of his predecessors who come from the same socioeconomic group as described by X_{0i} and I_i, have the same grade from the middle school and pick (hs, pt).

[52] As elsewhere in the model, we maintain the assumption that students do not target programs incompatible with their track, which implies $w_{ig} = 0.5$.

[53] Although one could derive the exact distribution of ε_{i0} given (**??**) and the definition of s_{it} in (4), we make a shortcut by approximating this distribution with $N[0, \sigma_{\varepsilon 0}(I_i, g_{0i})]$ in our estimation and simulation code.

Scores are valued by the students in so much as they improve chances of admission to selective colleges. By definition, the noise-free score at the first attempt $\bar{s}_{i1} = s_{i1}(hs_i, pt_i) - \varepsilon_{i1}$. Hence, at the end of middle school, the student chooses the level of pre-test investment (hs, pt) that maximizes his expected welfare conditional on the observables:

$$V_0(X_{0i}, g_{0i}, I_i, c_i) = \max_{hs,pt} \{E[V_1(s_{i1}(hs, pt) - \varepsilon_{i1}, \varepsilon_{i1}, I_i)|X_{0i}, g_{0i}, I_i] - c_i(hs, pt)\}, \qquad (16)$$

where $c_i = \{u_{i,hs}, v_{i,pt}, w_{i,hs,pt}\}_{hs,pt}$ denotes the full set of student-specific cost shocks for all values of hs and pt. Thus, equation (16) extends Bellman's equation (5) in the main model by adding an extra time period $t = 0$.

A.3.3 Estimation Results

We estimate parameters associated with schooling choices independently for the three income groups. We obtain the returns from schooling $\rho_{hs,pt,g_0,I}$ from equation (**??**) via OLS. As is common in such models, omitted ability bias may be affecting our results: for example, if better students choose prep schools, the estimates of returns from extra tutoring will be upward biased. However, controlling for the middle school GPA should address the problem, to the extent that the GPA captures the unobserved ability. Returns from attending Anatolian/science schools may still be biased as these schools use entrance exams to select students with highest ability.

Table 18 reports the estimates of schooling returns, $\rho_{hs,pt,g_0,I}$ in equation (**??**), relative to those from going to public school and taking no extra tutoring. The estimates are almost always significantly different from zero and their ranking is in line with common sense. Students from Anatolian and science schools outperform private school students, who in turn get higher scores than public school graduates. Extra tutoring in prep schools is also associated with higher first-time scores, irrespective of initial ability and high school choice. This is in line with anecdotal evidence that prep schools target knowledge specific to the entrance exam as even students going to selective Anatolian/Science schools seem to gain from prep schools.

Estimates of the costs, $\gamma_{hs,pt,g_0,I}$, are obtained by using equation (16). Given our distributional assumptions on idiosyncratic costs of schooling, this model of choice boils down to a random intercept logit (see Train (2009)). We substitute the estimates of the value function, V_1, obtained in step 3 of the estimation routine described in Section 4, and estimate the cost parameters using simulated maximum likelihood. Relative costs of various schooling options are reported in Table 18. The estimates are noisy, but the general pattern is clear: schooling options that involve more effort tend to be associated with higher gains and tend to be more costly. Given that, by its definition, the placement payoff is between zero and one, these schooling costs are very high. This is, again, in line with anecdotal evidence that high school students in Turkey who compete for seats in top colleges are under enormous pressure. Overall, choices made during the high school period

Income	Low			Middle			High		
Middle school GPA	A	B	C	A	B	C	A	B	C
Gains in score, $\rho_{hs,pt,g_0,I}$									
Anatolian	28.9	27.5	22.9	27.8	35.5	34.2	47.4	30.8	16.4
(no prep)	(2.9)	(2.7)	(4.6)	(3.4)	(3.5)	(9.4)	(5.4)	(8.8)	(6.5)
Anatolian	42.6	43.7	38.7	44.1	43.7	40	45.9	42.5	38.9
(prep)	(1.3)	(1.5)	(4.1)	(1.4)	(1.2)	(3.0)	(3.4)	(2.0)	(3.7)
Private,	11	12.1	21.3	5.21	15.8	30.8	16.9	10.6	16.9
(no prep)	(2.0)	(2.8)	(10.0)	(2.2)	(2.3)	(11.2)	(5.4)	(3.5)	(10.7)
Private,	25.2	27	34.3	23.3	24.8	16.3	25.6	25.4	26.1
(prep)	(1.4)	(1.8)	(6.5)	(1.5)	(1.3)	(6.6)	(3.5)	(2.1)	(4.1)
Public,	18	14.1	13.3	18.2	13.9	13.6	19.7	11.7	12.3
(prep)	(1.5)	(1.1)	(1.8)	(1.6)	(1.0)	(2.1)	(3.7)	(1.9)	(3.5)
Costs, $\gamma_{hs,pt,g_0,I}$ (significant at the 1% level — in **bold**)									
Anatolian	0.287	0.292	0.361	**0.229**	0.298	0.292	0.387	0.262	0.177
(no prep)	(2.07)	(3.01)	(5.50)	(0.08)	(0.17)	(0.22)	(0.18)	(1.09)	(1.92)
Anatolian	0.27	0.331	0.351	**0.296**	**0.306**	**0.281**	0.291	0.272	0.248
(prep)	(0.49)	(1.09)	(2.63)	(0.09)	(0.02)	(0.04)	(1.77)	(1.11)	(0.65)
Private,	0.121	0.197	0.316	0.050	0.143	0.267	0.133	0.088	0.146
(no prep)	(1.23)	(2.83)	(4.38)	(0.06)	(0.11)	(0.30)	(0.20)	(0.57)	(0.91)
Private,	0.159	0.234	0.382	0.144	**0.173**	0.148	0.138	0.143	0.151
(prep)	(0.21)	(1.58)	(4.69)	(0.07)	(0.03)	(0.16)	(1.37)	(0.77)	(0.42)
Public,	0.112	0.085	0.085	**0.115**	0.076	**0.075**	0.112	0.039	0.043
(prep)	(0.13)	(0.24)	(0.62)	(0.04)	(0.04)	(0.02)	(0.80)	(0.87)	(0.73)

Bootstrapped standard errors are in parentheses. Cost parameters significant at the 1% level are bolded. All gains and costs are relative to the baseline option: public school with no tutoring. Costs are rescaled to be in the same units as the placement payoff. Middle school GPA controls for the student's initial ability: A is the highest, C is the lowest.

Table 18. Schooling Choices: Gains and Costs.

Income group	Low		Middle		High	
Policy	Current	1 attempt	Current	1 attempt	Current	1 attempt
Anatolian, no prep	0.02	0.02	0.01	0.02	0.01	0.01
Anatolian, prep	0.21	0.25	0.35	0.48	0.52	0.61
Private, no prep	0.04	0.03	0.02	0.01	0.01	0.01
Private, prep	0.13	0.14	0.18	0.20	0.23	0.24
Public, no prep	0.31	0.26	0.14	0.05	0.04	0.01
Public, prep	0.26	0.27	0.26	0.21	0.17	0.11

Simulated shares of students who go to public schools with no tutoring, private schools with no tutoring, etc.
Policies: current – unlimited retaking, 1 attempt max.

Table 19. Simulated schooling choices: unlimited retaking vs 1 attempt max

have a much higher impact on placement scores and the costs incurred in the process than retaking decisions.

A.3.4 Gains from Banning Retaking

In the counterfactual experiments presented in Section 6, we assumed that the policy changes do not affect students' level of pre-exam investment before attempt 1 or afterwards. It is natural to look at the extent to which this affects our results. If restrictions on retaking result in a big increase in pre-exam investment in high school, the only effect of such a ban might be to move effort expended to a prior stage. Students may increasingly choose costly private schools over public ones and enroll in private tutoring.

To address these concerns, we re-run our simulations using the augmented model that explicitly accounts for high school choice. We only consider a complete ban on retaking in order to avoid issues potentially caused by endogenous effort between attempts.

In line with intuition, the retaking ban puts more pressure on students to perform well in the first attempt, so that fewer of them choose public schools. As shown in Table 19, the percentage of students who choose Anatolian schools and extra tutoring grows in all three income groups. As a result, the distribution of placement scores shifts to the right after one allows for endogenous schooling. Yet, this distribution is still dominated by the one in the unlimited retaking scenario, as illustrated in Figure 7. Consequently, admission cutoffs in the no-retaking scenario remain lower than in the unlimited retaking regime.

Finally, we obtain the expected gains from the ban for each expected first-time score decile and income group. Figure 8 depicts the point estimates and the respective 95 percent confidence intervals obtained via bootstrapping. The gains depicted here account for costs of pre-exam investment in high school, in contrast to those plotted in Figure 5b. The inverse U shaped gains are less pronounced and the average gain is halved due to welfare-reducing investment effects coming from the ban. The estimates are also noisier due to the high standard errors on the cost parameters.

Figure 7. Distribution of Placement Scores: Fixed vs Endogenous Schooling

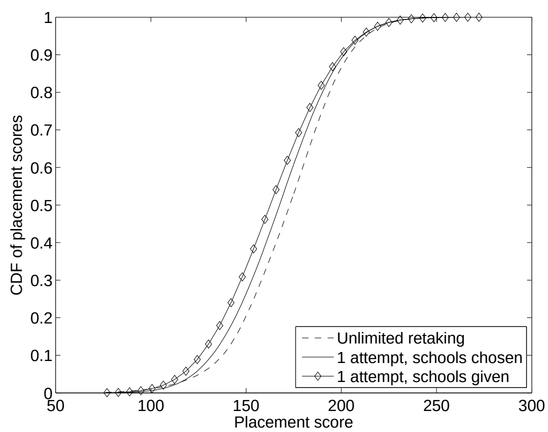
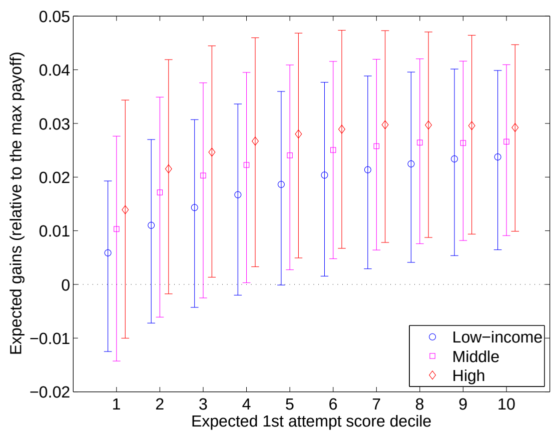


Figure 8. Gains from Banning Retaking if School Choices are Endogenous



Nevertheless, the results confirm our main finding: the vast majority of students would gain from the retaking ban, and this gain is significant, irrespective of their income.

A.4 Is Reducing Retaking Always Beneficial?

One might suspect, on the basis of the counterfactual exercises above that impediments to retaking should be unambiguously beneficial. This section shows that this is not the case and that welfare could go either way depending on the supply of seats and the distribution of student abilities.

There is a unit mass of applicants injected into the system every year. Student ability, θ_i, is drawn from a mixture of two normals: 45 percent of the population draw from $N[\mu = -4, \sigma = 1]$, while the rest draw from $N[4, 1]$. In other words, we have advantaged and disadvantaged students in the proportion of $55/45$. The score in attempt t is comes from $s_{it} = \theta_i + \varepsilon_{it}$, where the shock $\varepsilon_{it} \sim N[0, 4]$. For simplicity, we abstract from any learning between attempts. On the supply side, there is one college program with a mass of seats equal to 0.56. The payoff from being placed into this program is normalized to unity. The outside option is to drop out and earn zero. The students can retake at a cost of ψ. The future is discounted with $\delta = 0.9$. The equilibrium is characterized by the admission cutoff s^*. Those students whose scores are above s^* are placed into the program.

Those who are not, either retake or drop out. There is an ability threshold θ^*: students whose ability is above θ^* find it optimal to retake until placed.

We simulate the effect of raising costs ψ from 0.1 to 0.2. Figure 9a shows the change in ex ante expected welfare conditional on ability: the solid line depicts welfare under $\psi = 0.1$, while the dashed line corresponds to $\psi = 0.2$; the dotted vertical lines mark the cutoff abilities θ^*, while the shaded histograms show the ability distribution among first-time takers.

The increase in ψ discourages retaking: θ^* increases from -0.57 to 1.09 as depicted in Figure 9a by the vertical dashed lines. This creates an excess supply of seats; to restore equilibrium, the threshold s^* has to shift down from 4.31 to 4.15. Disadvantaged students benefit from this shift, as the probability of getting above the cutoff from the first attempt goes up. On the other hand, when they do retake they pay more, and this reduces their welfare. These two effects are very close for low ability students so that there is a small net improvement in their welfare. Advantaged students, on the other hand, are not as fortunate. While they benefit from the increased probability of placement, they have to pay more to retake, and the latter effect dominates.

There are slightly more seats than there are advantaged students who, along with the lucky low-ability agents who score high in the first attempt, are the ones who get in. A small reduction in the supply of seats from 0.56 to 0.55 puts pressure on the high types and raises the score cutoffs quite dramatically, from 4.31 to 6.34, as it triggers much more retaking among the high-ability students. The retaking surge occurs because there is a great deal to be gained from being placed and, as the standard deviation of ε_{it} is high, the high types tend to have a good chance of getting in.

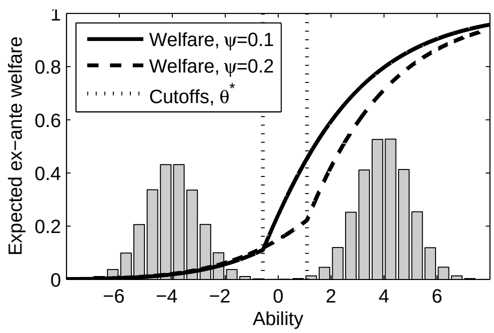
The results are quite sensitive to the parameters of the simulation. In this new scenario with $\psi = 0.55$, raising the cost of retaking sharply reduces the intensity of competition: the admission cutoff s^* falls from 6.34 to 4.91. As a result, all students benefit, especially the disadvantaged type as depicted in Figure 9b.

B Technical Details

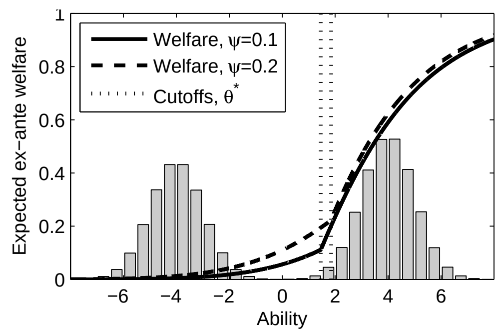
B.1 Misrepresentation of Preferences

While students in our model have no incentive to misrepresent their preferences, the actual allocation mechanism might provide such incentives. Consider the following example, borrowed from Balinski and Sönmez (1999). Students i and i' apply to two programs, r and r'. Admissions to r are based on Y-ÖSS-SAY scores, while program r' accepts Y-ÖSS-EA. Programs r and r' have one seat each, so that the number of applicants equals the number of seats. Students get the following pairs of Y-ÖSS-SAY and Y-ÖSS-EA scores: $s_i = (180, 190)$, $s_{i'} = (190, 180)$. Student i prefers r to r', while i' prefers r' to r.

Figure 9. Welfare Response to Increasing the Costs of Retaking



(a) Supply of seats =0.56



(b) Supply of seats = 0.55

If students report their preferences truthfully, they are allocated to their second choices: on the first iteration, the mechanism allocates r to the highest scorer in Y-ÖSS-SAY, and r' to the highest scorer in Y-ÖSS-EA. Suppose that i plays strategically and removes r' from his preference list. The first iteration of the mechanism will tentatively allocate r' to i' in the Y-ÖSS-EA category, as i' is the only student who claims this seat. Seat r will be allocated to i', too, as i' scores highest in Y-ÖSS-SAY. Then, the mechanism will account for i' preferences by placing him to r' and unblocking r. On the second iteration, r, the only remaining seat, will be allocated to i, the only remaining student.

Note that the conditions described above can only occur in very special (if not unlikely) circumstances. First, misrepresentation works because i believes that his actions affect admission cutoffs of programs targeted by his direct competitors. While this belief may be reasonable in a two-person game, it is hard to justify in a national placement system—an environment with hundreds of thousands of participants. Second, it is critical in the above example that student preferences are misaligned with their scores. A common strategy, however, is to prepare for the test according to one's preferences: students who prefer programs accepting Y-ÖSS-SAY to those accepting Y-ÖSS-EA devote more practice to math and natural sciences rather than language or social studies. Preferences for college subject also affect the choice of high school track, which, in turn, affects placement scores via the GPA weight. Thus, student scores are likely to reflect rather than contradict their preferences. Such examples are not possible in our setup, as we have a continuum of agents.

B.2 Equilibrium Conditions

In order to derive a system of equations for the retaking threshold functions $e_t(\bar{s}, I)$ and the inverse cutoff scores $r^*(s)$, it is useful to rewrite Bellman's equation (5) so that it recursively defines the continuation value of retaking for all attempts:

$$VC_t(\bar{s}_{it}, I_i) = \delta \int_{-\infty}^{\infty} \left\{ \left[\Phi\left(\frac{e_{t+1}(\bar{s}_{it} + \lambda_{it+1}, I_i)}{\sigma_\varepsilon(I_i)}\right) VC_{t+1}\left(\bar{s}_{it} + \lambda_{it+1}, I_i\right) \right.\right.$$
$$\left. + \int_{e_{t+1}(\bar{s}_{it}+\lambda_{it+1},I_i)}^{\infty} U(r^*(\bar{s}_{it} + \lambda_{it+1} + \varepsilon_{it+1}, I_i) \frac{1}{\sigma_\varepsilon(I_i)} \varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) d\varepsilon_{it+1} \right]$$
$$\left. \times \frac{1}{\sigma_{\lambda t+1}(I_i)} \varphi\left(\frac{\lambda_{it+1} - \mu_{\lambda t+1}(I_i)}{\sigma_{\lambda t+1}(I_i)}\right) \right\} d\lambda_{it+1} - \psi_t(I_i) \tag{17}$$

We use Φ and φ to denote the standard normal c.d.f. and density, respectively.

To find a solution to this recursive system, we use the assumption that for $t \geq 4$, $\lambda_{it} = 0$ and $\psi_t(I_i) = \psi_4(I_i)$. This implies that after the fourth attempt, the structure of future shocks and payoffs looks exactly the same for the student irrespective of t. Therefore, for all $t \geq 4$, $VC_t = VC_4$ and, as a result, $e_t = e_4$. This simplifies equation 17 so that it takes the following form:

$$VC_4(\bar{s}_{it}, I_i) = \delta \left[\Phi\left(\frac{e_4(\bar{s}_{it}, I_i)}{\sigma_\varepsilon(I_i)}\right) VC_4\left(\bar{s}_{it}, I_i\right) \right.$$
$$\left. + \int_{e_4(\bar{s}_{it},I_i)}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i) \varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) \frac{d\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)} \right] - \psi_4(I_i) \tag{18}$$

B.2.1 Solving for the Retaking Threshold Used After Attempt 3

One can solve equation (18) for VC_4:

$$VC_4(\bar{s}_{it}, I_i) = \left[1 - \delta\Phi\left(\frac{e_4(\bar{s}_{it}, I_i)}{\sigma_\varepsilon(I_i)}\right) \right]^{-1}$$
$$\times \left[\delta \int_{e_4(\bar{s}_{it},I_i)}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i) \varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) \frac{d\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)} - \psi_4(I_i) \right] \tag{19}$$

The above equation contains the optimal cutoff $e_4(\bar{s}_{it}, I_i)$, which is yet to be found. In order to characterize $e_4(\bar{s}_{it}, I_i)$ that solves Bellman's equation (5), it is convenient to use the following function:

$$W(e, \bar{s}_{it}, I_i) = U(r^*(\bar{s}_{it} + e); I_i) \left[1 - \delta\Phi\left(\frac{e}{\sigma_\varepsilon(I_i)}\right) \right]$$
$$- \delta \int_{e}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i) \varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) \frac{d\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)} + \psi_4(I_i) \tag{20}$$

Note that $W(e, \bar{s}_{it}, I_i)$ has the same sign as the difference between the placement payoff and the value of retaking given the retaking threshold of e. Thus, a finite e_4 solves the Bellman's equation for attempts $t \geq 4$, if it satisfies the following conditions:[54]

$$W(e_4, \bar{s}_{it}, I_i) = 0, \quad \text{and} \quad W(e, \bar{s}_{it}, I_i) < 0 \; \forall e < e_4. \tag{21}$$

The Bellman's equation may also have two corner solutions: $e_4(\bar{s}_{it}, I_i) = \infty$ (always retake) and $e_4(\bar{s}_{it}, I_i) = -\infty$ (never retake). The former strategy satisfies equation (5) if $W(e, \bar{s}_{it}, I_i) < 0$ for all values of e. The latter is optimal if $W(e, \bar{s}_{it}, I_i) > 0$ for any e.

Note that $W(e, \bar{s}_{it}, I_i)$ is monotonic in e. Let e' be greater than e. Then,

$$\begin{aligned}
W(e', \bar{s}_{it}, I_i) \geq\; & U(r^*(\bar{s}_{it} + e); I_i) \left[1 - \delta\Phi\left(\frac{e'}{\sigma_\varepsilon(I_i)}\right)\right] \\
& - \frac{\delta}{\sigma_\varepsilon(I_i)} \int_{e'}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i)\varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) d\varepsilon_{it+1} + \psi_4(I_i) \\
=\; & U(r^*(\bar{s}_{it} + e); I_i) \left[1 - \delta\Phi\left(\frac{e'}{\sigma_\varepsilon(I_i)}\right)\right] \\
& - \frac{\delta}{\sigma_\varepsilon(I_i)} \int_{e}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i)\varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) d\varepsilon_{it+1} + \psi_4(I_i) \\
& + \frac{\delta}{\sigma_\varepsilon(I_i)} \int_{e}^{e'} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i)\varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) d\varepsilon_{it+1} \\
\geq\; & U(r^*(\bar{s}_{it} + e); I_i) \left[1 - \delta\Phi\left(\frac{e'}{\sigma_\varepsilon(I_i)}\right) + \delta\Phi\left(\frac{e'}{\sigma_\varepsilon(I_i)}\right) - \delta\Phi\left(\frac{e}{\sigma_\varepsilon(I_i)}\right)\right] \\
& - \frac{\delta}{\sigma_\varepsilon(I_i)} \int_{e}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}), I_i)\varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) d\varepsilon_{it+1} + \psi_4(I_i) = W(e, \bar{s}_{it}, I_i).
\end{aligned}$$

Monotonicity implies that any solution to (21) is always unique.

If the costs of retaking are extremely high or low, the system (21) has no finite solutions. The former occurs if

$$\psi_4(I_i) > \frac{\delta}{\sigma_\varepsilon(I_i)} \int_{-\infty}^{\infty} U(r^*(\bar{s}_{it} + \varepsilon_{it+1}, I_i)\varphi\left(\frac{\varepsilon_{it+1}}{\sigma_\varepsilon(I_i)}\right) d\varepsilon_{it+1}.$$

In this case, $\lim_{e \to -\infty} W(e, \bar{s}_{it}, I_i) > 0$. Since W is non-decreasing in e, $W > 0$ for all values of e. That is, even the worst possible placement is preferred to retaking. A utility-maximizing agent would never retake: $e_4(\bar{s}_{it}, I_i) = -\infty$.

Another corner solution is to always retake. This is optimal if students derive utility from retaking instead of bearing a cost, or, more precisely, if $\psi_4(I_i) < -(1 - \delta)$. In this case, $W(e, \bar{s}_{it}, I_i) < 0$ for all values of e: the best possible placement gives a payoff of 1, while the value of retaking infinitely is $-\psi_4(I_i)/(1 - \delta) > 1$. Thus, the optimal strategy is $e_4(\bar{s}_{it}, I_i) = \infty$.

[54] We are implicitly assuming that the placement payoff is a continuous function of the score. Relaxing this assumption would require some extra technical twists in the proof.

B.2.2 Solving for the Retaking Thresholds Used in Attempts 1–3

After one finds the retaking threshold function $e_4(\bar{s}_{it}, I_i)$ for every I_i, the thresholds for attempts 1–3 can be obtained by backward induction.

First, given $e_4(\bar{s}_{it}, I_i)$, one can find the continuation value function, $VC_4(\bar{s}_{i4}, I_i)$, from equation (19). Knowing VC_4, one can use equation (17) to obtain $VC_3(\bar{s}_{i3}, I_i)$ as a function of $\bar{s}_{i3}$ and I_i.

Second, the optimal threshold e_3 is found from Bellman's equation (5). If $U(r^*(\bar{s}_{i3} + \varepsilon_{i3}); I_i) \geq VC_3(\bar{s}_{i3}, I_i)$ for all ε_{i3}, the optimal solution is to accept any placement: $e_3(\bar{s}_{i3}, I_i) = -\infty$. If $U(r^*(\bar{s}_{i3} + \varepsilon_{i3}); I_i) < VC_3(\bar{s}_{i3}, I_i)$ for any ε_{i3}, the student should retake irrespective of his score: $e_3(\bar{s}_{i3}, I_i) = \infty$. If neither is true, the retaking threshold is finite; it can be found from

$$U(r^*(\bar{s}_{it} + e_t(\bar{s}_{it}, I_i); I_i) = VC_t(\bar{s}_{it}, I_i), \quad \text{and} \tag{22}$$
$$U(r^*(\bar{s}_{it} + \varepsilon_{it}, I_i); I_i) < VC_t(\bar{s}_{it}, I_i) \text{ if } \varepsilon_{it} < e_t(\bar{s}_{it}, I_i)$$

Intuitively, students at the threshold should be exactly indifferent between retaking and being placed, while all students below should prefer retaking.

Functions e_2 and e_1 satisfying Bellman's equation are obtained recursively via the same steps as used above to find e_3. Since the solution for $e_4(\bar{s}_{it}, I_i)$ is unique given $r^*(s)$, e_1, e_2 and e_3 are unique by construction.

B.2.3 Solving for the Inverse Cutoff Function

By definition, $r^*(s_i)$ is an index of the best seat that the placement algorithm allocates to exam taker i if i chooses to stop retaking. Since the serial dictatorship algorithm allocates seats according to one's ranking in terms of score, $r^*(s_i)$ is the highest ranked seat unclaimed by exam takers who score above i. Note that, in equilibrium, no higher-scoring student would claim seats worse than $r^*(s_i)$, and no lower-scoring student would get seats above $r^*(s_i)$. Thus, $1 - r^*(s_i)$, the mass of seats ranked higher than $r^*(s_i)$, is equal to the mass of students who score higher than s_i and who choose placement over retaking.[55] Students who choose retaking over placement and score higher than i do not affect the value of $r^*(s_i)$: retaking implies that the student fails to be placed, either because he submits an empty list or because his preferred programs are not feasible with his score. For any of the above reasons, retakers do not occupy seats in the current year that student i could otherwise claim.

To express this more formally, let $N_t(I)$ denote the number of exam takers from the income group I in attempt t. $G_t(s, I)$ is the mass of students who belong to the income group I, are in attempt t, whose scores are greater than s, and who choose placement over retaking. According to

[55] In case the mass of higher ranked students is above unity (the mass of seats offered for placement), $r^*(s_i) = 0$: the outside option (indexed as $r = 0$) is in unlimited supply.

the above reasoning, the best seat that the score s_i can buy is

$$r^*(s_i) = \max\left\{0, 1 - \sum_{I=1}^{3}\sum_{t=1}^{\infty} G_t(s_i, I)\right\} \tag{23}$$

The number of first-time takers in each income group, $N_1(I)$, is exogenously given in the data; their total mass $\sum_{I=1}^{3} N_1(I)$ is normalized to unity. In order to derive a formula for $G_t(s, I)$, we integrate the density of t-time takers over their individual state variables, $\bar{s}_{it}$ and ε_{it}. We include only those students whose ε_{it} is above $\max\{e_t(\bar{s}_{it}, I_i), s - \bar{s}_{it}\}$ so that we count everyone who: i) fills a seat, and ii) gets a score higher than s:

$$\begin{aligned} G_t(s, I) &= N_t(I)\int_{-\infty}^{\infty}\left[\int_{\max\{s-\bar{s}_{it}, e_t(\bar{s}_{it}, I)\}}^{\infty} \varphi\left(\frac{\varepsilon_{it}}{\sigma_\varepsilon(I)}\right)\frac{d\varepsilon_{it}}{\sigma_\varepsilon(I)}\right] f_{\bar{s}_t|a_t,I}(\bar{s}_{it}, I)d\bar{s}_{it} \\ &= N_t(I)\int_{-\infty}^{\infty}\left[1 - \Phi\left(\frac{\max\{s - \bar{s}_{it}, e_t(\bar{s}_{it}, I)\}}{\sigma_\varepsilon(I)}\right)\right] f_{\bar{s}_t|a_t,I}(\bar{s}_{it}, I)d\bar{s}_{it} \end{aligned} \tag{24}$$

The above expression includes $f_{\bar{s}_t|a_t,I}(\bar{s}_{it}, I)$, a density of noise-free scores in attempt t among t-time takers[56] who belong to income group I. $N_t(I)$ and $f_{\bar{s}_t|a_t,I}(\bar{s}_{it}, I)$ have to satisfy the following conditions:

$$\begin{aligned} &N_{t+1}(I) = N_t(I) - \lim_{s\to-\infty} G_t(s, I), \\ &f_{\bar{s}_{t+1}|a_{t+1},I}(\bar{s}_{it+1}, I) = \frac{N_t(I)}{N_{t+1}(I)}\int_{-\infty}^{\infty} f_{\bar{s}_t|a_t,I}(\bar{s}_{it}, I)\Phi\left(\frac{e_t(\bar{s}_{it}, I)}{\sigma_\varepsilon(I)}\right)\varphi\left(\frac{\bar{s}_{it+1} - \bar{s}_{it} - \mu_{\lambda t+1}(I)}{\sigma_{\lambda t+1}(I)}\right)\frac{d\bar{s}_{it}}{\sigma_{\lambda t+1}(I)} \end{aligned} \tag{25}$$

The first condition states that the number of exam takers in attempt $t+1$ has to be equal to that in attempt t minus the mass of students who choose placement after t^{th} attempt. The second condition relates the density of noise-free scores among t and $(t+1)$-time takers. It accounts for attrition due placement (the term $\Phi\left(\frac{e_t(\bar{s}_{it}, I)}{\sigma_\varepsilon(I)}\right)\frac{N_t(I)}{N_{t+1}(I)}$) and the change in one's noise-free score due to the shock λ_{it+1} (the term $\varphi\left(\frac{\bar{s}_{it+1}-\bar{s}_{it}-\mu_{\lambda t+1}(I)}{\sigma_{\lambda t+1}(I)}\right)\sigma_{\lambda t+1}(I)^{-1}$). The integral accounts for the fact that a given $\bar{s}_{it+1}$ can be obtained from any value of $\bar{s}_{it}$ by adding a shock $\lambda_{it+1} = \bar{s}_{it+1} - \bar{s}_{it}$.

Equations (25) describe evolution of cohort's noise-free scores and size between attempts taking the initial conditions, $f_{\bar{s}_1|a_1,I}$ and $N_1(I)$, as given. The density of noise-free scores among first-time takers, $f_{\bar{s}_1|a_1,I}$, is determined from the GPA and the score equations (2) and (4). Condi-

[56] The subscript a_t denotes the event that the student does not choose placement until attempt t. That is, a_t occurs for student i if i's shocks satisfy $\varepsilon_{i\tau} < e_\tau(\bar{s}_{i\tau}, I_i)$ for all $\tau < t$.

tional on student characteristics, X_i and I_i, $\bar{s}_{i1}$ has normal distribution in the population:

$$\bar{s}_{i1}|X_i, I_i \sim N\left[\mu_{\bar{s}_1}(X_i, I_i), \sigma^2_{\bar{s}_1}(I_i)\right], \tag{26}$$
$$\mu_{\bar{s}_1}(X_i, I_i) = X_i'\left[\frac{\beta_g(I_i)}{2} + \beta(I_i)\right]$$
$$\sigma^2_{\bar{s}_1}(I_i) = \left[\frac{\alpha_g(I_i)}{2} + \alpha(I_i)\right]' \Sigma_\theta(I_i)\left[\frac{\alpha_g(I_i)}{2} + \alpha(I_i)\right] + \left(\frac{\sigma_{\varepsilon g}(I_i)}{2}\right)^2$$

Thus, the density $f_{\bar{s}_1|a_1,I}$ is a mixture of normal densities:

$$f_{\bar{s}_1|a_1,I}(\bar{s}_{i1}|I_i) = \int \frac{1}{\sigma_{\bar{s}_1}(I_i)}\varphi\left[\frac{\bar{s}_{i1} - \mu_{\bar{s}_1}(X_i, I_i)}{\sigma_{\bar{s}_1}(I_i)}\right] f_{X|I}(X_i|I_i)dX_i \tag{27}$$

where $f_{X|I}$ is the density of X_i conditional on i's income group.

B.2.4 The System of Steady State Conditions

The steady state equilibrium is fully characterized by the inverse cutoff function $r^*(s)$ defined in (23). In order for $r^*(s)$ to generate a steady state equilibrium, it has to satisfy the following conditions:

- The definition in (23) depends on the mass of exam takers who have scores above s and choose placement, $G_t(s, I)$, by attempt t and income group I. $G_t(s, I)$ is, in turn, defined in the system of equations (24) and (25). The initial conditions for (24–25) are given by (26) and (27).

- The numbers of placed students and their scores in (24–25) depend on student equilibrium behavior as given by $e_t(\bar{s}_{it}, I_i)$, the threshold for ε that triggers retaking.

  - For $t \geq 4$, $e_t(\bar{s}_{it}, I_i)$ is found by solving student's dynamic optimization problem using (20) and (21),
  - For $t < 4$, e_t is found from the same dynamic problem by backward induction using (17) to find the continuation value and (22) to obtain $e_t(\bar{s}_{it}, I_i)$.

  The threshold function e_t is uniquely defined given r^*.

- The total mass of placed students in all attempts, $\lim_{s\to-\infty}\sum_{t=1}^{\infty} G_t(s, I)$, should be equal to the number of fresh high school graduates, $N_1(I)$, for every income group, $I = 1, 2, 3$.

B.3 Comparative Statics

In this section, we simulate a simplified version of the model to develop some intuition. We assume that initial ability (i.e., noise-free placement score in the first attempt, $\bar{s}_{i1}$) is drawn from a normal

	Baseline	$Var[\varepsilon] \times 4$	$Var[\lambda] \times 4$	$\psi \times 2$	$U(r) = \sqrt{r}$	+10% top seats
$E(attempts)$	1.19	1.48	1.57	1.03	1.03	1.21
$E(V_1(\bar{s}_{i1}, \varepsilon_{i1}))$	0.32	0.29	0.29	0.33	0.66	0.41

Table 20. Simulated Comparative Statics

distribution, $N[130, 25]$, with the mean and the standard deviation close to those of the actual ÖSS-SAY score used to place students in the science track. In the baseline specification, we let $\delta = 0.9$, $\psi_t = 0.05$, the structural utility function take a constant relative risk aversion (CRRA) form, $U(r) = r^2$. The learning shocks are also normal with $E[\lambda_{it}] = 0$ and $Var[\lambda_{it}] = 150$ in attempts 2 through 4, and with no further learning in later attempts. We set the variance of noise in the score equation to be the same in all attempts and $Var[\varepsilon_{it}] = 25$.

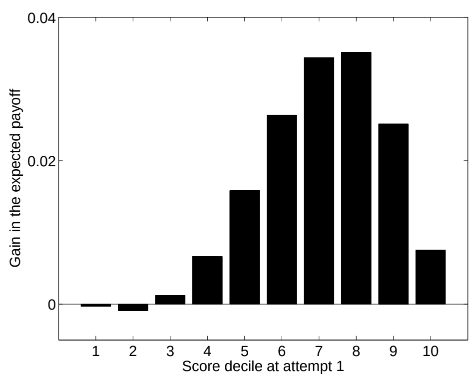
In columns 3 and 4 of Table 20, we quadruple the variance of ε_{it} and of λ_{it} respectively. We see that as this happens, the expected number of attempts rises, but the expected utility falls. The former makes sense as an increase in randomness makes people who fare badly in a given attempt more likely to retake. However, the negative externality retakers inflict on others makes expected utility fall when retaking rises. In column 5, we double retaking costs and this reduces retaking while raising expected utility. This suggests that policies that reduce retaking costs, such as more frequent exams, may be a bad idea. In column 6, we make agents risk averse rather than risk loving. As expected, this reduces the number of retakes. In column 7, we increase the number of seats at the top school by 10 percent. This increases the expected number of retakes as the prize from retaking becomes more accessible, and raises the expected utility. Note that the seemingly reasonable response of increasing seats as a response to a backlog of students might actually increase the backlog.

Can banning retaking reduce welfare under certain circumstances? Are the negative spillovers associated with retaking enough so as to have banning retaking raise expected welfare or welfare of most agents? If agents are homogeneous, then it can be shown[57] that banning retaking must raise welfare. But when agents are heterogeneous in terms of their initial ability, this result no longer holds for everyone. The simulations suggest that the opportunity to retake is valued by the majority when the placement payoffs induce love for risk. This occurs if the quality of seats is very far from uniform: seats at the top have much higher quality than those in the middle. In our baseline case, most students gain from banning retaking, as in Figure 10a below. In an alternative specification that reduces retaking costs and makes agents more risk loving, $\psi_t = 0.01$, $U(r) = r^8$, banning retaking results in the three highest score deciles gaining, but the majority loses as shown in Figure 10b. The direct effect of banning retaking is negative and more so for those who tend to retake

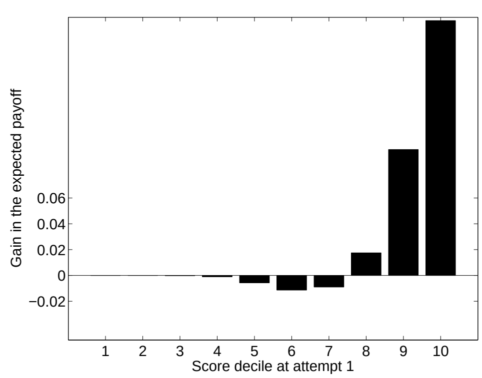
[57] The proof is available upon request.

more often. There is also a general equilibrium effect of banning retaking which is positive as competitive pressures are reduced. The probability of retaking falls with ability. Banning retaking insulates top students from competitive pressures and raises their welfare while reducing welfare for lower-ability students, who were more likely to retake. This illustrates the redistributional aspects of such a reform: the majority may in fact prefer unlimited retaking though the losses of the majority are less than the gains of the minority.[58]

Figure 10. Preventing Retaking: Welfare Consequences



(a) Baseline Case



(b) Risk Loving Agents

B.4 Estimating the Factor Model

In step one, we estimate the parameters of the GPA and the four subject score equations

$$g_i = X_i'\beta_g(I_i) + \theta_i'\alpha_g(I_i) + \varepsilon_{ig}, \qquad (28)$$
$$s_{ij1} = X_i'\beta_j(I_i) + \theta_i'\alpha_j(I_i) + \varepsilon_{ij1}, \quad j \in \{M, T, Sc, SS\}.$$

[58] If, in addition to heterogeneity in agents and schools, there are gains from matching better agents to better schools, retaking may help improve the match. In this environment, banning retaking can reduce aggregate welfare.

We obtain the estimates of $\beta_g(I_i)$ and $\beta_j(I_i)$ by running the respective outcome variable in (28) on X_i using a subsample of first-time takers in each of the three income groups: $I_i = 1, 2, 3$. By leaving only first-time takers in the estimation sample, we avoid issues caused by learning and self-selection into retaking.

Then, we use the residuals from the above regressions to pin down the factor loadings $\alpha_g(I_i)$ and $\alpha_j(I_i)$, the covariance matrix of θ, $\Sigma_\theta(I_i)$, and the standard deviations of the idiosyncratic shocks, $\sigma_{\varepsilon g}(I_i)$ and $\sigma_{\varepsilon j}(I_i)$, where j stands for math, Turkish, science and social studies. The residuals contain the effects of unobservables and random shocks that sum up to a total of seven factors, $(\theta_{iv}, \theta_{iq}, \varepsilon_{i0}, \varepsilon_{iM1}, \varepsilon_{iT1}, \varepsilon_{iSc1}, \varepsilon_{iSS1})$. Factor loadings capture a possibly differential effect that verbal and quantitative unobservable abilities may have on scores in different subjects. In order to identify the loadings and the distributions of all shocks we rely on the standard assumption from the literature on factor models that the idiosyncratic shocks ε and the vector of common factors θ are jointly independent in the population of first-time takers within each income group (assumption 2 in the main body).[59]

Given joint independence of shocks, one can identify $\alpha_g(I_i)$, $\alpha_j(I_i)$, the joint density of the common factors $[\theta_{iv}, \theta_{iq}]$, and the densities of the transitory shocks ε_{ig} and ε_{ij1} non-parametrically (see Freyberger (2012) for more details). However, non-parametric estimation of the above densities would make our steps 2 and 3 computationally formidable. By imposing joint normality on ε_{ig}, ε_{ij1} and θ_i, we circumvent this problem. Under normality, all of the above distributions are completely characterized by their variance-covariance structure, which we identify using moment conditions derived below.

Let R_i be a vector of the five residuals from the system of equations (28):

$$R_i = \begin{bmatrix} \alpha_{g1}(I_i) & \alpha_{g2}(I_i) \\ \alpha_{M1}(I_i) & \alpha_{M2}(I_i) \\ \alpha_{T1}(I_i) & \alpha_{T2}(I_i) \\ \alpha_{Sc1}(I_i) & \alpha_{Sc2}(I_i) \\ \alpha_{SS1}(I_i) & \alpha_{SS2}(I_i) \end{bmatrix} \begin{bmatrix} \theta_{iv} \\ \theta_{iq} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ig} \\ \varepsilon_{iM1} \\ \varepsilon_{iT1} \\ \varepsilon_{iSc1} \\ \varepsilon_{iSS1} \end{bmatrix} \tag{29}$$

Note that we can relabel common factors and the respective loadings in these equations as follows: let $\tilde{\theta}_{iq} = \alpha_{M1}\theta_{iv} + \alpha_{M2}\theta_{iq}$ and $\tilde{\theta}_{iv} = \alpha_{T1}\theta_{iv} + \alpha_{T2}\theta_{iq}$. We can invert these two equations so that θ_{iv} and θ_{iq} are expressed in terms of $\tilde{\theta}_{iv}$ and $\tilde{\theta}_{iq}$. Substituting for θ_{iv} and θ_{iq} in terms of $\tilde{\theta}_{iv}$ and $\tilde{\theta}_{iq}$ into the above system gives us a normalized set of equations where the coefficients on $\tilde{\theta}_{iv}$ and $\tilde{\theta}_{iq}$

[59] Note that the common factors θ_{iv} and θ_{iq} are allowed to correlate without any restrictions.

in the math and Turkish equations are $\tilde{\alpha}_M = [0, 1]$ and $\tilde{\alpha}_T = [1, 0]$ respectively:

$$R_i = \begin{bmatrix} \tilde{\alpha}_{g1}(I_i) & \alpha_{g2}(I_i) \\ 0 & 1 \\ 1 & 0 \\ \tilde{\alpha}_{Sc1}(I_i) & \tilde{\alpha}_{Sc2}(I_i) \\ \tilde{\alpha}_{SS1}(I_i) & \tilde{\alpha}_{SS2}(I_i) \end{bmatrix} \begin{bmatrix} \tilde{\theta}_{iv} \\ \tilde{\theta}_{iq} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ig} \\ \varepsilon_{iM1} \\ \varepsilon_{iT1} \\ \varepsilon_{iSc1} \\ \varepsilon_{iSS1} \end{bmatrix} \tag{30}$$

Since both (29) and (30) have the same residual R_i on the left-hand side, the models defined in (29) and (30) are observationally equivalent. Given that the choice of normalization does not affect the data generated by the model, we will impose $[\alpha_{M1}, \alpha_{M2}] = [0, 1]$ and $[\alpha_{T1}, \alpha_{T2}] = [1, 0]$.

Let $\alpha(I_i)$ denote the matrix on the right hand side of (30) that contains all factor loadings. The covariance matrix of R_i conditional on income can be expressed as

$$E[R_i R_i' | I_i = I] = \alpha(I)' \Sigma_\theta(I) \alpha(I) + \Sigma_\varepsilon(I).$$

$\Sigma_\varepsilon(I)$ is a diagonal matrix with the variances of ε's on the diagonal.

The left-hand side of the above equation is a 5×5 matrix; it can be estimated from the data by approximating the expectation with the mean. By symmetry, only 15 elements of this matrix need to be considered. On the right hand side, we have five variances of the ε shocks, two variances and one covariance of the common factors θ_i, and six unrestricted factor loadings $\alpha(I)$ for each income group. Therefore, we have 15 equations and 14 unknowns. We obtain the estimates of $\alpha(I)$, $\Sigma_\theta(I)$ and $\Sigma_\varepsilon(I)$ using GMM, with these equations as moment conditions.

B.5 Identification of Learning Shocks and Retaking Thresholds

Let $f_{y|x}$ denote the density of y conditional on x, $\bar{S}_{it} = [\bar{s}_{i1}, \ldots, \bar{s}_{it}]$ be the trajectory of i's noise-free scores up to attempt t, and a_{it} be a dummy for retaking until at least attempt t. The shock $\nu_{it} = \lambda_{it} + \varepsilon_{it}$ combines the effects of learning and noise in scores. To streamline the argument, we abstract away from the observable demographics X_i and income I_i; one way to interpret our derivations here is to treat them as conditional on X_i and I_i.

The following equation is instrumental in proving that the decision rules and the distribution of learning shocks are identified:

$$\begin{aligned} f_{s_{t+1},g|a_{t+1}=1}(s_{it+1},g_i) &= \int f_{s_{t+1},\bar{S}_t,g|a_{t+1}=1}(s_{it+1},\bar{s}_{it},g_i)d\bar{S}_{it} \\ &= \int f_{\nu_{t+1},\bar{S}_t,g|a_{t+1}=1}(s_{it+1}-\bar{s}_{it},\bar{S}_{it},g_i)d\bar{S}_{it} \\ &= \int f_{\bar{S}_t,g|a_{t+1}=1}(\bar{S}_{it},g_i)f_{\nu_{t+1}}(s_{it+1}-\bar{s}_{it})d\bar{S}_{it} \\ &= \int \frac{\Pr\{a_{it+1}=1|\bar{S}_{it},g_i\}f_{\bar{S}_t,g}(\bar{S}_{it},g_i)}{\Pr\{a_{it+1}=1\}}f_{\nu_{t+1}}(s_{it+1}-\bar{s}_{it})d\bar{S}_{it} \\ &= \frac{1}{\Pr\{a_{it+1}=1\}}\int\prod_{\tau=1}^{t}\left[\Phi\left(\frac{e_\tau(\bar{s}_{i\tau})}{\sigma_\varepsilon}\right)\right]f_{\bar{S}_t,g}(\bar{S}_{it},g_i)f_{\nu_{t+1}}(s_{it+1}-\bar{s}_{it})d\bar{S}_{it} \end{aligned} \tag{31}$$

The first step in this derivation is almost tautological: we integrate the density of $(s_{it+1}, \bar{S}_{it}, g_i)$ over $\bar{S}_{it}$ and obtain the density of (s_{it+1}, g_i). In the second step, we change variables using the fact that $s_{it+1} = \bar{s}_{it} + \lambda_{it+1} + \varepsilon_{it+1}$. The third line exploits independence of $\nu_{it+1} = \lambda_{it+1} + \varepsilon_{it+1}$ of any shocks that trigger retaking in attempt t ($a_{it+1} = 1$) as well as of the GPA and the noise-free score of student i. In line four, we apply Bayes' theorem to transform the conditional density of $(\bar{s}_{it}, g_i)$. Finally, in the last step we compute the probability that in all t attempts student's $\varepsilon_{i\tau}$ was below the retaking threshold $e_\tau(\bar{s}_{it})$.

We build our identification argument recursively. Suppose that one knows the "initial" density of noise-free scores and GPA, $f_{\bar{s}_1,g}(\cdot,\cdot)$, the retaking threshold functions and the parameters of f_λ up to attempt t: $\{e_\tau(\cdot), \mu_{\lambda\tau+1}, \sigma_{\lambda\tau+1}\}_{\tau=1}^{t-1}$. The objective is to identify $e_t(\cdot)$, $\mu_{\lambda t+1}$ and $\sigma_{\lambda t+1}$.

B.5.1 Identifying Selection into Retaking

As demonstrated schematically in Section 4, the retaking thresholds are backed out from the way GPA density evolves between attempts. Thus, we will concentrate on GPA density by integrating

s_{it+1} out of (31):

$$f_{g|a_{t+1}=1}(g_i) = \frac{1}{\Pr\{a_{it+1}=1\}} \iint \prod_{\tau=1}^{t} \left[\Phi\left(\frac{e_\tau(\bar{s}_{i\tau})}{\sigma_\varepsilon}\right)\right] f_{\bar{S}_t,g}(\bar{S}_{it}, g_i) f_{\nu_{t+1}}(s_{it+1} - \bar{s}_{it}) d\bar{S}_{it} ds_{it+1}$$
$$= \frac{1}{\Pr\{a_{it+1}=1\}} \int \prod_{\tau=1}^{t} \left[\Phi\left(\frac{e_\tau(\bar{s}_{i\tau})}{\sigma_\varepsilon}\right)\right] f_{\bar{S}_t,g}(\bar{S}_{it}, g_i) d\bar{S}_{it}$$
$$= \frac{1}{\Pr\{a_{it+1}=1\}} \int \underbrace{\left\{\int \prod_{\tau=1}^{t} \left[\Phi\left(\frac{e_\tau(\bar{s}_{i\tau})}{\sigma_\varepsilon}\right)\right] f_{\bar{s}_t \ldots \bar{s}_2|\bar{s}_1}(\bar{s}_{it}, \ldots, \bar{s}_{i1}) d\bar{s}_{it} \ldots d\bar{s}_{i2}\right\}}_{H_t(\bar{s}_{i1})}$$
$$\times f_{\bar{s}_1,g}(\bar{s}_{i1}, g_i) d\bar{s}_{i1} \tag{32}$$

In the last step, we use the fact that the distribution of s_{it} does not depend on g_i after conditioning on s_{i1}.

First, consider $t = 1$. In the above formula, $H_1(\bar{s}_{i1}) = \Phi\left(\frac{e_1(\bar{s}_{i1})}{\sigma_\varepsilon}\right)$ is an unknown function. The density of GPA among second-time takers, $f_{g|a_2=1}(g_i)$, and the probability of surviving $t+1$ attempts, $\Pr\{a_{it+1} = 1\}$, can be non-parametrically identified using the data. The density $f_{\bar{s}_1,g}(\bar{s}_{i1}, g_i)$ is known from step 1 of the estimation algorithm. Thus, in order to find an unknown $H_1(\bar{s}_{i1})$, we have to solve a Fredholm integral equation of the first kind with the kernel given by $f_{\bar{s}_1,g}(\bar{s}_{i1}, g_i)$.

In general, solutions to such integral equations do not necessarily exist, depending on the structure of the kernel function. In our case, however, one can use conditional normality of $\bar{s}_{i1}$ and g_i to demonstrate existence and uniqueness. The joint normal[60] density $f_{\bar{s}_1,g}(\bar{s}_{i1}, g_i)$ can always be rewritten as $f_{\bar{s}_1,g}(\bar{s}_{i1}, g_i) = \xi_1(\bar{s}_{i1} - \kappa g_i)\eta_1(\kappa g_i)$, where the coefficient κ and the functions $\xi_1(\cdot)$ and $\eta_1(\cdot)$ are known. Let $G_i = \kappa g_i$ and $\zeta_1(G_i) = \frac{f_{g|a_2=1}(G_i/\kappa)\Pr\{a_{i2}=1\}}{\eta_1(G_i)}$; note that the function $\zeta_1(G_i)$ is known, too. Then, the above equation takes the following form:

$$\zeta_1(G_i) = \int_{-\infty}^{\infty} H_1(\bar{s}_{i1})\xi_1(\bar{s}_{i1} - G_i) d\bar{s}_{i1} \tag{33}$$

As shown in Polyanin and Manzhirov (2008), this equation can be solved for $H_1(\cdot)$ using a deconvolution technique based on Fourier transforms of the kernel and the left hand side, $\xi_1(\cdot)$ and $\zeta_1(\cdot)$. Knowing $H_1(\cdot)$, one can easily find the retaking threshold: $e_1(\bar{s}_{i1}) = \sigma_\varepsilon \Phi^{-1}(H_1(\bar{s}_{i1}))$.[61]

Suppose now that $t = 2$. That is, we know the distributions of all shocks leading to attempt 2 and the retaking threshold function for the first attempt. Our objective is to find $e_2(\cdot)$. Note that

[60] Recall that every density in this part of the appendix is conditional on the demographics and income. Conditional on X_i and I_i, the scores and the GPA are jointly normal.
[61] Note that this argument falls apart if the score and the GPA are not correlated, that is, if $\kappa = 1$.

we can follow the same set of steps as above and find $H_2(\cdot)$ given our knowledge of $f_{g|a_3=1}(\cdot)$, $\Pr\{a_{i3}=1\}$ and $f_{\bar{s}_1,g}(\cdot,\cdot)$. That is, we know the function on the left-hand side of

$$H_2(\bar{s}_{i1}) = \Phi\left(\frac{e_1(\bar{s}_{i1})}{\sigma_\varepsilon}\right)\int \Phi\left(\frac{e_2(\bar{s}_{i2})}{\sigma_\varepsilon}\right) f_{\bar{s}_2|\bar{s}_1}(\bar{s}_{i2},\bar{s}_{i1})d\bar{s}_{i2} \tag{34}$$

Again, we have a Fredholm integral equation of the first kind. The unknown is $\Phi\left(\frac{e_2(\bar{s}_{i2})}{\sigma_\varepsilon}\right)$, while the kernel $f_{\bar{s}_2|\bar{s}_1}(\bar{s}_{i2},\bar{s}_{i1})$ is a normal density, which can be written down as a (known) function of $\bar{s}_{i2}-\bar{s}_{i1}$: $f_{\bar{s}_2|\bar{s}_1}(\bar{s}_{i2},\bar{s}_{i1}) = f_{\lambda 1}(\bar{s}_{i2}-\bar{s}_{i1})$. It immediately follows that equation (34) takes a form reminiscent of equation (33):

$$\zeta_2(\bar{s}_{i1}) = \frac{H_2(\bar{s}_{i1})}{\Phi\left(\frac{e_1(\bar{s}_{i1})}{\sigma_\varepsilon}\right)} = \int_{-\infty}^{\infty}\Phi\left(\frac{e_2(\bar{s}_{i2})}{\sigma_\varepsilon}\right) f_{\lambda 1}(\bar{s}_{i2}-\bar{s}_{i1})d\bar{s}_{i2}$$

One can find $\Phi\left(\frac{e_2(\bar{s}_{i2})}{\sigma_\varepsilon}\right)$ (and hence $e_2(\cdot)$) by applying the same deconvolution technique as above taking $\zeta_2(\cdot)$ and $f_{\lambda 1}(\cdot)$ as known inputs.

This argument is easily extended to the values of $t>2$.

B.5.2 Identifying the Distribution of Learning Shocks

Identification of $\mu_{\lambda t+1}$ and $\sigma_{\lambda t+1}$, the parameters of learning shocks, is also based on equation (31). It is convenient to rewrite this equation by integrating out the trajectories $\bar{S}_{it-1}$ and the GPA g_i:

$$\begin{aligned}
f_{s_{t+1}|a_{t+1}=1}(s_{it+1}) &= \frac{1}{\Pr\{a_{it+1}=1\}}\iint \prod_{\tau=1}^{t}\left[\Phi\left(\frac{e_\tau(\bar{s}_{i\tau})}{\sigma_\varepsilon}\right)\right] f_{\bar{S}_t,g}(\bar{S}_{it},g_i) f_{\nu_{t+1}}(s_{it+1}-\bar{s}_{it})d\bar{S}_{it}dg_i \\
&= \frac{1}{\Pr\{a_{it+1}=1\}}\int R_t(\bar{s}_{it}) f_{\nu_{t+1}}(s_{it+1}-\bar{s}_{it})d\bar{s}_{it} \qquad (35)\\
\text{where } R_t(\bar{s}_{it}) &= \iint \prod_{\tau=1}^{t}\left[\Phi\left(\frac{e_\tau(\bar{s}_{i\tau})}{\sigma_\varepsilon}\right)\right] f_{\bar{s}_t,\bar{S}_{t-1},g}(\bar{s}_{it},\bar{S}_{it},g_i)d\bar{S}_{it-1}dg_i
\end{aligned}$$

The function $R_t(\bar{s}_{it})$ captures the mass of applicants who survive $t+1$ attempts and whose trajectories of noise-free scores arrive to $\bar{s}_{it}$ by attempt t. We know the retaking thresholds up to attempt t: $e_1(\cdot),\ldots,e_t(\cdot)$. We also know the distributions of all shocks leading to attempt t, which implies that we know the joint density of GPAs and noise-free score trajectories $f_{\bar{s}_t,\bar{S}_{t-1},g}(\cdot,\cdot,\cdot)$. Therefore, the function $R_t(\bar{s}_{it})$ is known.

The density of scores in attempt $t+1$ can be nonparametrically identified from the data using scores of $t+1$-time takers. The probability of surviving t attempts, $\Pr\{a_{it+1}=1\}$, can be easily obtained from the data, too (e.g., by dividing the number of $t+1$-time takers by the number of first-time takers).

Thus, the only unknown in the integral equation (35) is the density $f_{\nu_{t+1}}(\cdot)$. In principle, it can be identified using nonparametric deconvolution. However, given our assumptions, ν_{it+1} is drawn from $N[\mu_{\lambda t+1}, \sigma^2_{\lambda t+1} + \sigma^2_{\varepsilon}]$. Thus, we only need to find the mean and the variance of ν_{t+1}, which we can use to obtain $\mu_{\lambda t+1}$ and $\sigma^2_{\lambda t+1}$, the main parameters of interest.

B.6 Moment Conditions for Steps 2 and 3

In what follows, we derive moment equations (8) and (12). Let t and t' be two attempt numbers such that $t \geq t'$. Then,

$$
\begin{aligned}
&\Pr[(g_i, s_{it+1}) \in \Omega_k, a_{it+1} = 1 | a_{it'} = 1, I_i = I] \qquad (36)\\
&= \int \prod_{\tau=t'}^{t} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{S}_{it}, X_i, I, a_{it'} = 1\} \\
&\quad \times f_{\bar{S}_t, g, X | I, a_{t'}=1}(\bar{S}_{it}, g_i, X_i, I) d\bar{S}_{it} dg_i dX_i \\
&= \int f_{\bar{S}_t, g, X | I, a_{t'}=1}(\bar{S}_{it}, g_i, X_i, I) \prod_{\tau=t'}^{t} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}, I\} d\bar{S}_{it} dg_i dX_i \\
&= \int f_{\bar{S}_t, s_{t'}, g, X | I, a_{t'}=1}(\bar{S}_{it}, s_{it'}, g_i, X_i, I) \prod_{\tau=t'}^{t} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}, I\} \\
&\quad \times d\bar{S}_{it} dg_i dX_i ds_{it'} \\
&= \int \left[\int f_{\bar{S}_t | s_{t'}, g, X, I, a_{t'}=1}(\bar{S}_{it}, s_{it'}, g_i, X_i, I) \prod_{\tau=t'}^{t} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}, I\} d\bar{S}_{it}\right] \\
&\quad \times dG(s_{it'}, g_i, X_i | I, a_{it'} = 1) \\
&= \int \left[\int \frac{\Pr\{a_{it'} = 1 | \bar{S}_{it}, s_{it'}, g_i, X_i, I\} f_{\bar{S}_t | s_{t'}, g, X, I}(\bar{S}_{it}, s_{it'}, g_i, X_i, I)}{\Pr\{a_{it'} = 1 | s_{it'}, g_i, X_i, I\}} \prod_{\tau=t'}^{t} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \right. \\
&\quad \left. \times \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}, I\} d\bar{S}_{it}\right] dG(s_{it'}, g_i, X_i | I, a_{it'} = 1) \\
&= \int \left[\int \prod_{\tau=1}^{t} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{it}, I\} \right. \\
&\quad \left. \times dF(\bar{S}_{it} | s_{it'}, g_i, X_i, I)\right] \frac{dG(s_{it'}, g_i, X_i | I, a_{it'} = 1)}{\Pr\{a_{it'} = 1 | s_{it'}, g_i, X_i, I\}}
\end{aligned}
$$

First, the left hand side of (8) is equal to the expected value of the joint probability of the following events: $\varepsilon_{i\tau} < e_\tau(\bar{s}_{i\tau}, I)$ for all $\tau = t', \ldots, t$ and the sum $\lambda_{it+1} + \varepsilon_{it+1}$ taking student i from $(g_i, \bar{s}_{it})$ into Ω_k. Given $\bar{S}_{it}$ and I, these events are independent; thus, their joint probability equals the product of the marginal probabilities. We take the expected value with respect of $(\bar{S}_{it}, g_i, X_i)$ —

noise-free score trajectories, GPA and socioeconomic characteristics. Note that the probability of the last event depends neither on X_i nor on $[\bar{s}_{i1} \dots \bar{s}_{it-1}]$. We use this to obtain the fourth line.

In line 5, we express the density of $(\bar{S}_{it}, g_i, X_i)$ via that of $(\bar{S}_{it}, s_{it'}, g_i, X_i)$. In line 7, we change the order of integration and split the joint density of $(\bar{S}_{it}, s_{it'}, g_i, X_i)$ into the conditional density of $\bar{S}_{it}$ and G, the measure of observables from attempt t' in the population of t'-time takers. Next, we express the density of $\bar{S}_{it}$ conditional on survival (a complex object that depends on prior retaking) via that in the population of first-time takers (a joint normal density). We do so by applying Bayes' theorem in line 9. Finally, we regroup terms and use the fact that $a_{it'} = 1$ holds if and only if $\varepsilon_{i\tau} < e_\tau(\bar{s}_{i\tau}, I_i),\ \tau = 1, \dots, t' - 1$ to calculate the probability of survival in the numerator. One can obtain moment conditions (8) from (36) by setting $t' = t$.

In order to derive (12), note that the noise-free score $\bar{s}_{it}$ stops evolving after $t = 4$. Thus, for $t \geq 4$ and $t' = 4$, the condition in (36) becomes

$$\Pr[(g_i, s_{it+1}) \in \Omega_k, a_{it+1} = 1 | a_{i4} = 1, I_i = I] = \int \left\{ \int \prod_{\tau=1}^{3} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \left[\Phi\left(\frac{e_4(\bar{s}_{i4}, I)}{\sigma_\varepsilon(I)}\right)\right]^{t-3} \times \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{i4}, I\} dF(\bar{S}_{i4} | s_{i4}, g_i, X_i, I) \right\} \frac{dG(s_{i4}, g_i, X_i | I, a_{i4} = 1)}{\Pr\{a_{i4} = 1 | s_{i4}, g_i, X_i, I\}}$$

Since $\bar{s}_{i\tau} = \bar{s}_{i4},\ \forall \tau \geq 4$, integrating over the tails $[\bar{s}_{i5}, \dots, \bar{s}_{it}]$ of noise-free score trajectories is redundant; therefore we only integrate over $\bar{S}_{i4}$ in the above formula.

We then take a sum of both sides in the above equation over $t \geq 4$ and obtain (12):

$$\sum_{t \geq 4} \Pr[(g_i, s_{it+1}) \in \Omega_k, a_{it+1} = 1 | a_{i4} = 1, I_i = I] = \int \left[\int \prod_{\tau=1}^{3} \Phi\left(\frac{e_\tau(\bar{s}_{i\tau}, I)}{\sigma_\varepsilon(I)}\right) \frac{\Phi\left(\frac{e_4(\bar{s}_{i4}, I)}{\sigma_\varepsilon(I)}\right)}{1 - \Phi\left(\frac{e_4(\bar{s}_{i4}, I)}{\sigma_\varepsilon(I)}\right)} \times \Pr\{(g_i, s_{it+1}) \in \Omega_k | g_i, \bar{s}_{i4}, I\} dF(\bar{S}_{i4} | s_{i4}, g_i, X_i, I) \right] \frac{dG(s_{i4}, g_i, X_i | I, a_{i4} = 1)}{\Pr\{a_{i4} = 1 | s_{i4}, g_i, X_i, I\}}.$$

B.7 Simulation Algorithm

In our counterfactual policy experiments we simulate retaking behavior in the steady states that arise after policy changes. As shown in Section B.2, a steady state can be characterized by the inverse cutoff function $r^*(s)$. This function is a general equilibrium object; it changes whenever the policy environment is being changed. Among other things, it captures the level of competition in the exam. Given $r^*(s)$, each student knows what seat his score would buy, should he decide to be placed.

The objective of the simulation algorithm is to find $r^*(s)$, approximated with a piecewise-linear function on a uniform grid of 36 points: $s_{(1)}, \dots, s_{(36)}$. We define a mapping $T(r^*) : \mathbb{R}^{36} \to \mathbb{R}^{36}$ that takes the values $r^* = [r^*(s_{(1)}), \dots, r^*(s_{(36)})]$ as an argument:

1. Given a candidate r^*, the algorithm solves each student's dynamic decision problem, as described in Sections B.2.1 and B.2.2.
2. The solution to the dynamic problem is used to simulate student scores and retaking decisions.
3. Scores of exam takers are used to compute the inverse cutoffs $\widetilde{r}(s, r^*)$ that are consistent with student decisions. We use equation (23) for this purpose, substituting the mass of students $G_t(s, I)$ with its finite sample analog—the share of students who come from income group I, score above s and get placed in attempt t.
4. The mapping is given by

$$T(r^*) = \begin{bmatrix} r^*(s_{(1)}) - \widetilde{r}^*(s_{(1)}, r^*) \\ \dots \\ r^*(s_{(36)}) - \widetilde{r}^*(s_{(36)}, r^*) \end{bmatrix}$$

If r^* characterizes a steady state, $T(r^*) = 0$. We solve the system $T(r^*) = 0$ numerically and obtain r^*. Given r^*, we find student actions and payoffs and use them to compute the outcomes of interest in our counterfactual policy exercises.