

NBER WORKING PAPER SERIES

RETAKE IN HIGH STAKES EXAMS:
IS LESS MORE?

Kala Krishna
Sergey Lychagin
Verónica Frisancho Robles

Working Paper 21640
<http://www.nber.org/papers/w21640>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2015

We thank Pelin Akyol, Peter Kondor and Cemile Yavas for insightful discussions, as well as seminar participants at the Cardiff University, Central European University, University of Exeter, Pompeu Fabra University, University of St. Gallen, Stockholm University and the University of Warwick. Krishna is grateful to the Department of Economics at New York University for support in 2013-14 as a Visiting Professor. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Kala Krishna, Sergey Lychagin, and Verónica Frisancho Robles. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Retaking in High Stakes Exams: Is Less More?

Kala Krishna, Sergey Lychagin, and Verónica Frisancho Robles

NBER Working Paper No. 21640

October 2015

JEL No. C35,I23

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. We find that limiting retaking, though individually harmful given the equilibrium, actually increases expected welfare across the board. This result comes from a general equilibrium effect: retakers crowd the market and impose negative spillovers on others by raising acceptance cutoffs.

Kala Krishna
Department of Economics
523 Kern Graduate Building
The Pennsylvania State University
University Park, PA 16802
and NBER
kmk4@psu.edu

Verónica Frisancho Robles
Research Department
Inter-American Development Bank (IADB)
1300 New York Ave. NW
Washington, DC 20577
vfrisancho@iadb.org

Sergey Lychagin
Department of Economics
Central European University
Nador u. 9
Budapest 1051
Hungary
lychagins@ceu.hu

Retaking in High Stakes Exams: Is Less More?*

Kala Krishna

The Pennsylvania State University, CES-Ifo and NBER

Sergey Lychagin

Veronica Frisancho Robles

Central European University

IADB

6th October 2015

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. We find that limiting retaking, though individually harmful given the equilibrium, actually increases expected welfare across the board. This result comes from a general equilibrium effect: retakers crowd the market and impose negative spillovers on others by raising acceptance cutoffs.

In much of the world, both now and in the past, competitive exams have been used to select the best and brightest. The imperial examination required to be chosen as a civil servant in Imperial China is a classic example. Civil service exams remain common in many countries including China, Japan, India, the UK and the US. Admission to university in many countries is also similarly structured and is fiercely competitive. Students often

*Frisancho: Inter-American Development Bank, Research Department, 1300 New York Ave. NW, Washington, DC 20577 (e-mail: vfrisancho@iadb.org). Krishna: Kern Graduate Building, Room 523, The Pennsylvania State University, University Park, PA, 16802, USA, (e-mail: kmk4@psu.edu). Lychagin: Central European University, Nador u. 9, Budapest 1051, Hungary (e-mail: lychagins@ceu.hu). We thank Pelin Akyol, Peter Kondor and Cemile Yavas for insightful discussions, as well as seminar participants at the Cardiff University, Central European University, University of Exeter, Pompeu Fabra University, University of St. Gallen, Stockholm University and the University of Warwick. Krishna is grateful to the Department of Economics at New York University for support in 2013-14 as a Visiting Professor.

retake these exams multiple times; they spend enormous amounts of time, money and effort trying to improve their performance and get 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 spend 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. Suicides have been reported among students who learn the right answer for a question they missed. Despite such extreme duress, twenty percent retake in hopes of doing better.¹ In China, the infamous “gaokao” taken to enter university creates extreme stress. Shocking pictures of students hooked up to intravenous drips or taking oxygen while studying made the rounds.² Similar stories abound in other countries and settings.

Retaking has both positive and negative elements: on the plus side, retaking reduces the impact of bad luck on outcomes as it allows for second chances and so insures against downside risk. Those students who do poorly at the exam relative to what they expect would select into retaking, which could reduce the extent of student-college mismatch. Retaking 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 benefits an individual gains from moving up in the rankings necessarily comes at the cost of others.³ Moreover, admission tends to have rents associated with it as higher education is subsidized in much of the world. These rents are dissipated through excessive effort and retaking. Retakers increase competition for a given number of slots which has general equilibrium consequences: admission standards rise with more students competing for seats.

In the US, SAT exams nowadays are studied for quite intensely, and taken multiple times by students, especially by better off students. Moreover, private coaching for the SATs has also become the norm, especially among the well off. Our results suggest that this unrestricted opportunity to retake could have adverse consequences. Retaking is also related to the issue of “red-shirting⁴” in the US. In the US, children, especially boys, often start school a year late in the hope that this will allow them to do better than their peers. Deming and Dynarski (2008) provide a lucid summary of work on this topic.

Given the prevalence of retaking it is surprising that, at least to our knowledge, there

¹See 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.

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

³Due to this externality, retaking is likely to be excessive. In other words, such a contest has a zero sum nature as in Akerlof’s (1976) rat race. One might think that studying is good and, in fact, much of the contest literature tries to elicit more effort on the part of agents. However, the negative spillovers suggest that when effort is exerted only to improve standing, it is socially costly.

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

is no systematic analysis of its costs and benefits in a general equilibrium setting. The only previous 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 the exam where students choose whether or not to retake. Students are forward looking and weigh the expected benefits in terms of their future score with the costs of retaking. Intuitively, students who do worse than they should are the ones who retake. The model lets us answer the following questions that are key for policy: if retaking was limited, or even eliminated, what would be the consequences? Who would gain and who would lose? What is the effect on mismatch? 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% of them were at least in their 4th attempt. Though there are roughly as many seats as there are high school graduates, the large fraction of retakers creates an overhang. An increase in the number of seats, which may appear as an obvious way of clearing this overhang, will not solve the problem as retaking is an equilibrium phenomenon.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 methods of estimating dynamic models. In particular, identifying selection into retaking and improvement in scores between attempts is especially non-trivial in our case. To separate selection from learning we use the fact that High School GPA is unaffected by learning. Therefore, the distribution of HS GPA of retakers reflects only selection. On the other hand, exam performance of retakers has both selection and learning. By looking at the joint distribution of the two across attempts we can tease out learning from selection. 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 tends to under estimate learning.

We find that more advantaged students tend to have lower costs of retaking. Utility differs only for the best placements where the poor seem to value better schools far more than the rich. Learning gains are between 0.2 and 0.5 standard deviations of the placement

⁵See Hatakenaka (2006) for more on the challenges for the Turkish higher education system.

score and are highest in the middle income group and lowest among the rich.

An advantage of modelling equilibrium and estimating the structural parameters of the model as done here is that we can perform counterfactual experiments with the aim to guide policy. In steady state, we find that if retaking is prevented, most students tend to gain. 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. In our simulations, this general equilibrium effect dominates, so that everyone is *ex-ante* better off by restricting retaking. Nevertheless, if students are naive and cannot anticipate general equilibrium effects of such reforms, they will resist the restrictions.⁶

While our model captures the essence of the issues we choose to focus on, several limitations need to be pointed out at this stage. First, one of the benefits of retaking would be that a better match is obtained when second chances are given to students. In this paper, we do not postulate any gains from assortative matching since we have no way to identify them in our data. We do, however, attempt to capture part of this by looking at the extent to which students are under placed without retaking. Second, our estimated model accounts for endogenous effort but in a limited way. We model effort and costs incurred in high school by allowing students to choose between 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 captured by the fixed costs of retaking. 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. 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. 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) *per se* other than those that operate via placement. We do this because we have no data on the extent of such benefits. Fourth, in our estimation we assume that preferences are purely vertical, though they can differ by income class. While this assumption captures what seems to be a clear hierarchy between schools, it assumes away idiosyncratic preferences across majors within the science track and in terms of geographic location, for example.

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 US. They *calibrate a partial* equilibrium model and show that the practice of using the best SAT score serves to discrim-

⁶This is another example of the fallacy of composition at work.

inate 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 in their paper, 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*.

Another paper that looks at retaking and learning is Tornkvist and Henriksson (2004). It uses data on SweSAT, the Swedish version of the SAT, and documents patterns in it. Like in the US, 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. Their work is more descriptive than analytical. 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 use in a cross-sectional dataset such as ours 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 asks how to elicit more effort from agents as effort is what the principal cares about.⁷ 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.⁸

In what follows, we first lay out the data and a simple model that captures the essential aspects of the Turkish system. In Section 3 we discuss the intuition behind the model's identification and our estimation procedure. We report the estimation results in Section 4. Section 5 contains the counterfactual exercises. Section 6 concludes.

⁷For instance, see Fu (2007) and Fain (2009).

⁸In 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 like 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.

1 The Data

Turkey has a highly centralized college admission procedure. All potential college applicants in a given year have to take the ÖSS, Student Selection Exam, which is used for college placement and simultaneously administered all over the country once a year by OSYM (Student Selection and Placement Center). This exam attracts a great deal of attention and is considered a rite of passage for fresh high school graduates, irrespective of their plans to pursue a college education.

The exam is composed of multiple choice questions with negative marking for incorrect answers. Students' performance is evaluated in four subjects: Mathematics, Turkish, Science, and Social Studies. These *subject* scores together with the normalized high school GPA are used to construct the *placement* score. As students are encouraged to stay in their chosen tracks, those from the non science track applying to science programs are penalized in this process. Depending on the college program chosen by the student, different weights are applied to the four subjects tested in the exam resulting in placement scores that vary by program for a given student. However, over 82% of the students placed in 4 year programs from the Science track are placed using the score called ÖSS-SAY.⁹ For this reason, we focus on this score below.

After taking the placement exam and learning the results, the students submit their college preferences.¹⁰ In addition to their scores, students receive a booklet with previous year's cut-off scores for each program (i.e. the score of the last student admitted). Cut-off scores in the most popular programs are very stable across years. Placement is merit based: a student is placed in his most preferred program, conditional on the availability of seats after all the applicants with higher scores are placed.

Students fail to be placed if they are not eligible to put down preferences (i.e., their score is too low) or if all the choices they put down on their list are unavailable to them (i.e., they are filled up by better students). These students have the option of retaking the exam with no penalties but their current (not highest) score is used for placement. Students who are placed are also allowed to retake, but their placement score is penalized if they retake the following year. Given that competition for seats in good colleges is very intense, even a small penalty is enough to hurt their placement a lot. Only 6 percent of the current placements are from students already in 4 year colleges. In what follows, we remove enrolled applicants from the data and assume that one cannot apply to other programs after being placed.¹¹

Our data covers a random sample of about 42,731 students who took the ÖSS in 2002 and

⁹For more on how these scores differ from each other see Frisancho et. al (2013).

¹⁰Only those students who obtain more than a certain score are eligible to submit preferences.

¹¹Had we not assumed placement was terminal, we would have complicated the model a lot.

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%) across different data sources, and dropping those who retake while already enrolled in a university program (13%), as well as those with missing data (8%) we lose roughly 25% of the observations. We restrict 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 outcomes (raw scores, weighted scores, and placement outcomes). Since we want to measure high school performance across schools, we construct quality normalized GPAs (normalizing GPAs by school performance in the university entrance exam) to control for quality heterogeneity and grade inflation across high schools (see the Appendix for details).¹²

1.1 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% of the science track applicants were repeat takers.¹³ According to our data, approximately 80% of retakers are not employed at the time of the exam.

High retaking rates could arise from three sources: a low cost of retaking, a high value of a better placement, and a probable improvement in scores due to learning and uncertainty in test results. If costs of retaking are low, one would expect more retaking to occur. If there is randomness in the test results, and payoffs in terms of placements are convex, then it may well be worthwhile to retake as doing a bit better moves the student to a much more valued school.

How prevalent is retaking in different socioeconomic groups? Frisancho et. al (2013) suggests that the disadvantaged have greater learning gains than the advantaged. Consequently, we would expect them to retake more often. On the other hand, if the disadvantaged have higher costs of retaking, then they will be less likely to retake.

¹²It is worth noting that very few papers have explored the Turkish data set. Tansel (2005) studies the determinants of attendance at private tutoring centers and its effects on performance. Saygin (2011) 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.

¹³These numbers are much higher in the social studies track. Overall, about 67 percent are retakers.

Income	Low	Medium	High	Income	Low	Medium	High
Father's education				Funds for college			
Primary or less	6,567	4,305	1,320	Family/rental inc/scholarship	3,044	5,164	4,096
Middle/High school	2,538	4,702	2,392	Work	2,633	2,937	1,894
2-year higher education	136	1,014	812	Loan	3,837	4,127	2,002
College/Master/Phd	245	1,958	3,463	Other	682	657	481
Missing	710	906	486	High school city population			
Mother's education				<10k	1,345	971	263
Primary or less	8,658	8,656	2,702	10-50k	1,754	1,793	766
Middle/High school	1,009	2,921	2,756	50-250k	2,487	2,913	1,602
2-year higher education	26	307	973	250-1000k	1,637	2,284	1,374
College/Master/Phd	32	285	1,615	>1million	2,289	4,192	4,108
Missing	471	716	427	Missing	684	732	360
Father's occupation				Family size			
Employer	42	299	847	1-2 children	4,504	8,446	6,999
Works for wages/salary	3,847	8,474	5,569	>3 children	5,692	4,439	1,474
Self-employed	3,506	3,085	1,705	Gender			
Unemployed/not in LF	2,801	1,027	352	Female	4,185	5,761	3,921
Mother's occupation				Male	6,011	7,124	4,552
Employer	4	23	104	High school category			
Works for wages/salary	304	1,456	3,249	Public	6,951	7,262	3,326
Self-employed	534	355	331	Private	1,447	2,480	2,028
Unemployed/not in LF	9,354	11,051	4,789	Anatolian/Science	1,301	2,696	2,973
Expenditures on prep. schools				Other	497	447	146
No prep school	2,350	1,603	423	Access to internet			
Scholarship	410	395	245	Yes, at home	296	1,316	2,698
Less than 1b	3,790	5,211	2,137	Yes, not at home	2,733	4,455	2,911
1-2 b	906	2,936	2,790	No	6,792	6,681	2,666
More than 2b	219	776	2,203	Missing	375	433	198
Missing	2,521	1,964	675				

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 1: Demographic Composition of Exam Takers from the Science Track

The value of retaking will also depend on the marginal return from obtaining a higher score, for all socioeconomic backgrounds: if the value of doing better rises sharply at a particular score level, then students whose scores are close to that level would gain more from retaking and so retake more often. For example, if being placed at the worst school is much better than not being placed, students just below this school’s cut-off score should tend to retake more often.

Table 2 shows the number of students in low, middle and high income groups for each retaking attempt.¹⁴ The data suggests that the poor are more present among the multiple time takers, and more so for later attempts. 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 a lot at lower scores. In either case, this preliminary evidence suggests that eliminating or restricting retaking may affect the poor more adversely than other groups.

Income	Low	Medium	High
# of students, by attempt			
1	4,454	6,388	4,757
2	2,635	3,221	2,043
3	1,547	1,681	917
4	861	929	401
5+	699	666	355
Mean first-time score	126	132	140
Std.dev. of first-time score	23	23	23

Table 2: Number of Exam Takers by Attempt and Income

1.2 Learning and Selection into Retaking

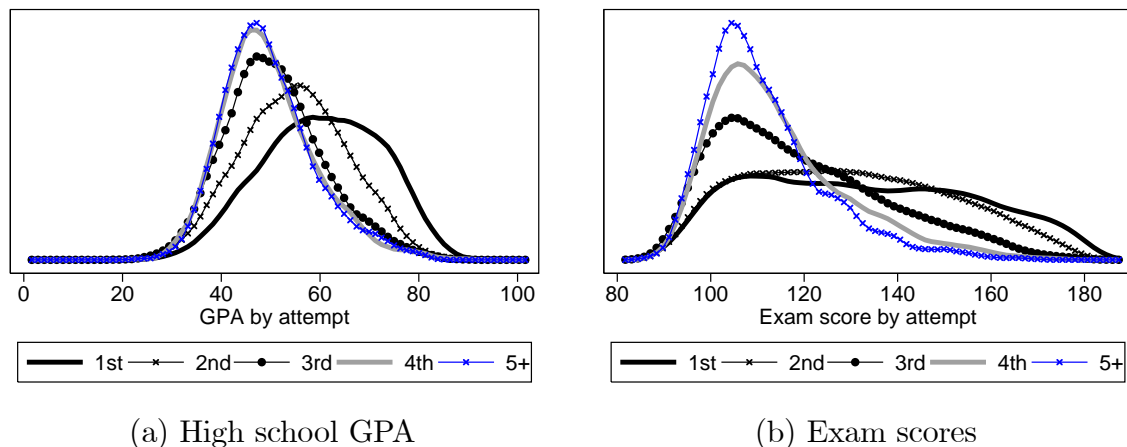
In Section 3, we show how students’ ability affect retaking rates and how much scores improve between attempts using the distributions of high school GPA and exam scores. Before we do so, we take a look at the raw data on both performance measures.

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

¹⁴Our 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 is reported) and there is an incentive to under report incomes as scholarship levels are related to income. We expect the order to be more correct than the level reported and this is why we use this coarse grouping.

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 worse students selecting into retaking (movement of the distribution to the left) dominating learning (movement of the distribution to the right).

Figure 1: Distributions of exam scores and high school GPA by attempt



The numbers by themselves say little about the desirability of allowing unlimited retaking or the gains from restricting it. To say anything in this context, we need to develop a model of retaking that clearly lays out the costs and benefits involved. We need to estimate this model’s parameters and use it to predict changes in student welfare in response to restrictions on retaking. This is what we turn to next.

2 Modelling the Turkish System

We model retaking decisions in an optimal stopping rule framework. We make the following key assumptions: i) students know their own ability though this is unobserved by the econometrician, ii) repeat takers may improve their score by taking a draw from a distribution that is allowed to vary with observables, and iii) performance in high school and at the entrance exam is partly determinate, coming from observables and unobserved ability, and partly random. We take a factor approach where the factors are the random performance shocks and the unobserved ability. In our model, ability will drive the correlation between high school grade point average (GPA) and raw verbal and quantitative exam scores, once the effect of observables is netted out.

After setting up our baseline model of retaking, we add on a prior stage that incorporates the choice of high school type and private tutoring. In this manner, we can incorporate the

effects of banning retaking on effort prior to taking the exam for the first time as banning retaking could intensify the rat race in high school.

2.1 Modelling Performance

There is a mass of infinitesimally small students. Each student has a high school GPA. As these may reflect differential grading practices across schools, GPAs are normalized to be comparable using the school’s performance in the university entrance exam. Details on how this is done are in the Appendix. We will postulate that the normalized high school GPA for a given student¹⁵ is given by

$$g = X'\beta_g + \theta'\alpha_g + \varepsilon_0. \quad (1)$$

where X is a vector of individual characteristics (laid out in Table 1) that do not vary in time and are potentially correlated with the student’s ability. The remaining terms, $\theta'\alpha_g + \varepsilon_0$, constitute the residual; $\theta = [\theta_q, \theta_v]'$ 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_q and θ_v , 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 is observed by the student, but unobserved by the econometrician. The shock, ε_0 , captures the randomness associated with the GPA. The distributions of θ and ε_0 could depend on X , but are required to be independent from each other conditional on X .

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

$$s_{jt} = X'\beta_j + \theta'\alpha_j + \sum_{\tau=2}^t \lambda_{j\tau} + \varepsilon_{jt} \quad (2)$$

where ε_{jt} is the corresponding error term for the student’s score in subject j (social studies, science, Turkish and math) and attempt t .¹⁶ We assume that the ε_{jt} ’s are i.i.d. conditional on X . $\lambda_{j\tau}$ denotes the student’s draw of the learning shock. The learning shocks are assumed to be independent over time though their distribution is allowed to depend on X and to vary across attempts. Moreover, conditional on X , the θ ’s, λ ’s and ε ’s are independent from each other. Note that learning shocks on the first attempt are by assumption zero as the exam is taken for the first time at the end of high school.

Factor loadings, α_g and α_j , do not vary across students in a group or attempts. They do

¹⁵We dispense with individual subscripts for ease of notation.

¹⁶Note that θ and ε_{jt} differ in that the draw of θ is the same for a given student for the GPA as well as exam scores, while the draws of the ε_0 and ε_{jt} ’s varies across them.

vary across income groups. The loadings in the math and Turkish equations are represented by $\alpha_m = [1, 0]'$ and $\alpha_T = [0, 1]'$ respectively: in other words, quantitative ability affects the math score but not the Turkish score and vice versa for verbal ability.¹⁷

The placement score in the t^{th} attempt, s_t , is a weighted sum of subject scores in that same attempt and the high school GPA using weights w that are publicly observed.

$$\begin{aligned}
s_t &= w_g g + \sum_j w_j s_{jt} \\
&= w_g g + X' \sum_j w_j \beta_j + \theta' \sum_j w_j \alpha_j + \sum_j \sum_{\tau=2}^t w_j \lambda_{j\tau} + \sum_j w_j \varepsilon_{jt} \\
&= w_g g + X' \beta + \theta' \alpha + \sum_{\tau=2}^t \lambda_{\tau} + \varepsilon_t \tag{3}
\end{aligned}$$

where the aggregate transitory shock to the t^{th} placement score is denoted as ε_t and λ_t captures the (permanent) learning shocks in the t^{th} attempt.¹⁸ We will abbreviate this to

$$s_t = \bar{s}_t + \varepsilon_t$$

where \bar{s}_t is the permanent component of the placement score at attempt t .

We assume that before taking the exam for the first time a student knows his X , θ , and ε_0 . Since the exam is taken for the first time at the end of high school and almost all high school graduates take it, we assume that there are no costs of taking the exam for the first time. Upon receiving his exam score, the student learns his ε_1 . Students put down their preferences only after knowing their score in the exam. Knowing his X , θ , ε_0 and ε_1 , the student decides whether or not to retake the exam. If the student does not retake, he accepts his most preferred outcome, which could be the option of not being placed. A student who decides not to retake cannot change his decision later.

There is a retaking cost which is incurred upon deciding to retake. This captures the fact that the exam is given only once a year and that preparation is costly. Second time takers study for the exam and so learn their λ_2 . Upon learning their score in the second attempt, they observe their ε_2 . A similar time line occurs for later attempts. The future is discounted at a common rate of δ by all students.

¹⁷This normalization is without loss of generality. See the Appendix for details.

¹⁸In what follows, we assume that the variance of ε_t is unaffected by t . However, note that the variance of ε_t and ε_0 can differ. This makes sense: the GPA is accumulated over the year so it may have a smaller transitory shock than the exam score.

2.2 Preferences and Utility Maximization

Admission decisions in Turkey are based on the placement score, s_t . As a result, those with better scores will have more options available to them, which yields them higher utility. The utility of accepting a placement with score s is denoted by $u(s, X)$. As scores define allocations in equilibrium, the utility derived from a given s comes from being allocated to the best seat that score allows. We let the utility vary by income group: for example, rich students may value the best schools only a little bit more than the next tier as their future is less dependent on the school they go to than that of a middle class student. As a result, some social groups may be more determined to compete for seats in top programs than others.

More formally, let $U(r; X)$ be the utility of having rank r . A higher rank is associated with a higher score and a better placement.¹⁹ We assume there is a continuum of seats and students so that there are no strategic elements involved. We assume that preferences are identical across all agents and purely vertical. That is, all students agree on which is the best school.²⁰ Though this is a strong assumption, it may be less objectionable in the Turkish context as there seems to be a clear hierarchy of schools, at least within tiers. The top tier includes the best public universities and private schools while bottom tier schools are those offering only two year programs and distance education. These assumptions provide us with a natural setting to get started as modelling preferences as well as dynamics will complicate things substantially.²¹

Though we assume all students have the same ordinal utility, the cardinal valuations, $U(r; X)$, are allowed to vary across students with different characteristics, X , i.e., income levels. Given $U(r; X)$ and the distribution of scores of non-retakers, $G(s)$, one can derive $u(s, X)$ as²²:

$$u(s, X) = U(G(s); X). \quad (4)$$

Note that the worst student has a rank of 0 while the best one has a rank of 1.²³ We normalize the utilities obtained by the worst and best students to be zero ($U(0, X) = 0$) and unity ($U(1, X) = 1$), respectively.

The distribution of scores, $G(s)$, is an equilibrium outcome. A change in the rules of

¹⁹Note that even in more general settings where preferences are not strictly vertical, the indirect utility is increasing in r as a higher rank allows for more options.

²⁰We choose not to specify a richer structure with preference heterogeneity and placement into different programs since our focus here is on retaking.

²¹For example, if students retake based on both preference and performance shocks, retaking choices would be much harder to estimate.

²²The mass of students who are placed is normalized to one in the steady state.

²³Being placed to the seat of the lowest rank may be interpreted as dropping out.

retaking will affect $G(s)$ and thus $u(s, X)$. If, for example, we are in the current system with unlimited, albeit costly, retaking and there are many students taking the exam at a given point of time, then a score of s may get one a middling rank and a mid level placement. However, if retaking is banned, then the number of students taking the exam will be much lower and score improvements through learning will be ruled out. In this case, the same score s may yield a far better rank and seat than the one obtained under the current system.

From equations (3), and (4) and the assumptions made above, the student's well-being in any attempt is entirely determined by the permanent component of the score, \bar{s}_t , and the corresponding transitory shock, ε_t . The student maximizes his utility by solving a dynamic optimization problem. Let $V_t(\bar{s}_t, \varepsilon_t, X)$ be the value function for attempt t and $VC_t(\bar{s}_t, X)$ be the continuation payoff. As usual:

$$V_t(\bar{s}_t, \varepsilon_t, X) = \max\{u(\bar{s}_t + \varepsilon_t, X), VC_t(\bar{s}_t, X)\} \quad (5)$$

where

$$VC_t(\bar{s}_t, X) = \delta E_{\varepsilon_{t+1}, \lambda_{t+1}|X} [V_{t+1}(\bar{s}_t + \lambda_{t+1}, \varepsilon_{t+1}, X)] - \psi_t(X).$$

$\psi_t(X)$ denotes retaking costs, which may vary with observables (such as income group as in our main econometric specification) and the subscript in $E_{\varepsilon_{t+1}, \lambda_{t+1}|X}$ emphasizes that the distributions of ε and λ can vary with X .

As defined above, \bar{s}_t is the permanent component of the exam score. As the student's utility is non-decreasing in his score, he is better off retaking when ε_t is below a threshold. Thus, a student's decision follows a simple rule: retake after the t^{th} attempt if

$$\varepsilon_t < e_t(\bar{s}_t, X) \quad (6)$$

That is, if the student's score is well below his predictions (based on innate ability and accumulated learning), he perceives his result as driven by bad luck and is likely to retake, expecting to do better next time.

In our estimation, we obtain $u(s, X)$, the utility function of being placed with score s in the existing equilibrium. Score s places the student in a seat with rank r , which is the best available seat after all students with a score above s are placed. Thus, $r = G(s)$. Consequently, the utility of obtaining a score s , $u(s, X)$, is identical to the utility of being placed in seat r , denoted by $U(r, X)$. Since we know the distribution of scores of those placed, $G(s)$, we can back out $U(r, X)$ from $u(s, X)$.

At this point, it is worth laying out the definition of the equilibrium. This is a Nash equilibrium in an anonymous game. Students have a common utility function $U(r, X)$. Each

student makes retaking decisions rationally (i.e. equation (5) holds) taking as given the level of competition as captured by $G(s)$, the distribution of scores of those placed. In turn, $G(s)$ is the equilibrium outcome consistent with such behavior. The system is in steady state if $G(s)$ does not change over time since the number of students exiting has to be equal to the number of students entering the system. Thus, the number of people graduating from high school (all of whom take the exam) must equal the sum of those who are placed in a university and those who choose to quit (take a seat at the bottom of the ranking). Out of steady state, equilibrium consists of a path for $G(s)$ and mass of agents over time, which is what agents expect when they are making their decisions. In its turn, this path is generated by the students' behavior. Note that $U(r, X)$ is the primitive utility, while $u(s, X)$ is an equilibrium outcome.

2.3 Schooling Choices Prior to the First Attempt

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 enrol into fee-paying high schools and pay for private tutoring after retaking is banned. To explore this possibility, we augment our model with a stage that describes the choice of high school type and additional tutoring.

Every student chooses between three broad categories of high schools: public, private and Anatolian/science. The student also makes a decision whether to get private tutoring for the entrance exam. In total, each student faces six options: public school with no tutoring, public school with tutoring, private school with no tutoring, and so on.

Schooling choices are associated with costs, which capture the fees and effort of keeping up with a more demanding curriculum. If the student chooses high school category h ($h = public, private, A/S$) and prep school tutoring p ($p = 0, 1$), he faces the cost of $c_{hp} = \bar{c}_{hp}(m, I) + c_p + c_h - \omega_{hp}$. The mean cost, $\bar{c}_{hp}(m, I)$, depends on the middle-school GPA m , which takes three values in the data, and income group, I . Student-specific shocks associated with school types and tutoring, c_p and c_h , are assumed to be jointly normal with an unknown covariance matrix Σ . Public school with no tutoring is set as the baseline option so that $\bar{c}_{pub,0} = c_0 = c_{pub} = 0$. Idiosyncratic shocks ω_{hp} are independently drawn from the type-1 Gumbel distribution.

Choosing a more expensive school type can affect performance in the college exam as each schooling option influences the student's expected score. The returns from schooling are also allowed to depend on the student's income. We assume that after conditioning on income, the expected improvement does not vary across students. Thus, from the perspective of the

	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(utility)	0.32	0.29	0.29	0.33	0.66	0.41

Table 3: Simulated Comparative Statics

student in middle school deciding on effort, the 1st-time choice-specific placement score (for a given income group) is

$$s_{1,hp} = X_0' \beta_0 + \Sigma_{h,p} \chi_{hp} d_{hp} + \varepsilon_{-1} \quad (7)$$

where d_{hp} is the set of dummies for the six schooling options and X_{0i} are the controls observed by the student in the middle school (family size, own gender, parent education, etc. but not high school type). The shock ε_{-1} reflects the uncertainty about the future score; its variance is allowed to depend on student's income and middle-school GPA.

Scores are valued by the students in so much as they improve chances of admission to selective colleges. By definition, the permanent component of the score at the first attempt, $\bar{s}_1 = s_{1,hp} - \varepsilon_1$. Hence, at the end of middle school, the student maximizes the following objective function

$$\max_{h,p} \left\{ a(I) E_{\varepsilon_{-1} | X_0, m, I} [V_1(s_{1,hp} - \varepsilon_1, \varepsilon_1)] - \bar{c}_{hp}(m, I) - c_p - c_h + \omega_{hp} \right\} \quad (8)$$

with respect to h and p , where V_1 is the attempt 1 value function defined in equation (5). Parameter a depends on income and captures the importance of expected placement payoffs relative to the costs incurred in high school.

2.4 Some 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}_1) is drawn from a normal 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 $\delta = 0.9$, $\psi = 0.05$, the structural utility function take a constant relative risk aversion (CRRA) form, $U(r) = r^2$. The learning shocks are also normal with $E[\lambda] = 0$ and $Var[\lambda] = 150$ 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 $Var[\varepsilon] = 25$.

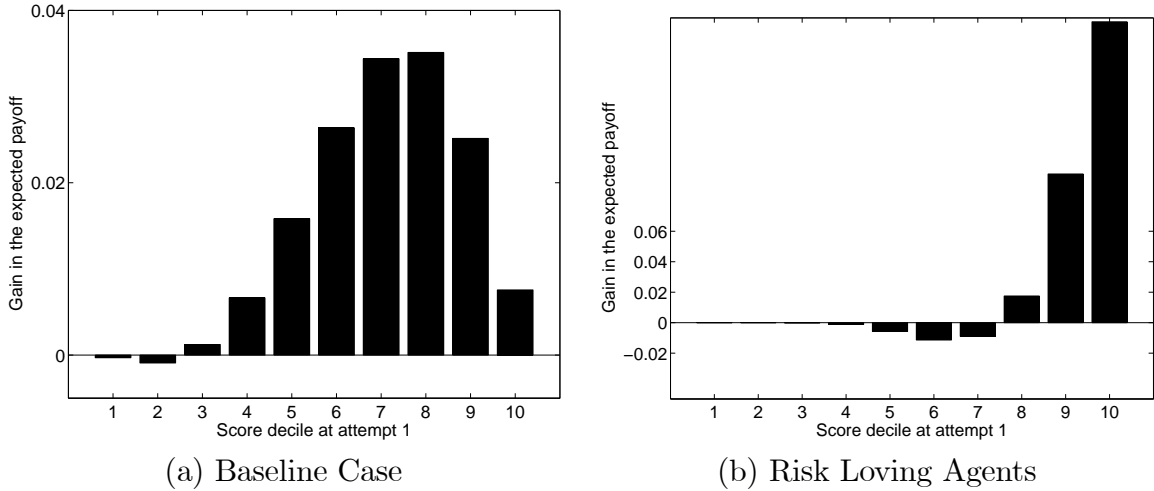
In columns 3 and 4 of Table 3, we quadruple the variance of ε and of λ respectively. We see that as this happens, the expected number of attempts rises, but expected utility falls.

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%. 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 (see Krishna, Lychagin, and Tarasov, 2015) that banning retaking must raise welfare. But when agents are heterogeneous in terms of their initial ability, this result no longer holds. The simulations suggest that risk loving agents will tend to want to retake. Banning retaking should result in losses for them, but gains for the more risk averse agents who do not want to retake and bear the burden of the negative externality inflicted by retakers. In our baseline case, most students gain from banning retaking as in Figure 2a below. In an alternative specification that reduces retaking costs and makes agents more risk loving, $\psi = 0.01$, $U(r) = r^8$, retaking is even more attractive. In this case, banning retaking results in the three highest score deciles gaining, but the majority loses as shown in Figure 2b. The direct effect of banning retaking is negative and more so for those who tend to retake 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 redistributive 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.²⁴

²⁴If, 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.

Figure 2: Preventing Retaking: Welfare Consequences



3 Identification

Our goal is to estimate the key structural parameters of the model by income group. We estimate the model from the viewpoint of someone taking the exam for the first time. There are three steps. In step 1, we use standard techniques from the literature on factor models to obtain the distributions of shocks to the high school GPA and to the scores as well as the distribution of unobserved innate abilities (denoted by f_{ε_0} , f_{ε_t} , and f_{θ} , respectively). In this stage, we also obtain factor loadings and the coefficients on X in the GPA and score equations given by (1) and (3). For simplicity, we assume that f_{θ} , f_{ε_0} , and $f_{\varepsilon_t} \forall t$ are normal which makes their estimation straightforward.²⁵ In step 2, we estimate the selection cutoffs defined in equation (6) and the distributions of the learning shocks, λ_t . Step 3 deals with the dynamic component; in this step, we estimate the costs of retaking, ψ , and the utility function defined in equation (4). We discuss the intuition behind each step below.²⁶

Step 1:

As practically every high school senior takes the university entrance exam, the sub-sample of first time takers is free of selection. By imposing normality on the distributions of ε_0 , $\varepsilon_t \forall t$, and θ , we can easily estimate the distributions of $\varepsilon_0, \varepsilon_t \forall t$, and θ as well as $\beta_g, \beta_j \forall j, \alpha_g$, and $\alpha_j \forall j$ from the five-equation system defined by (1) and (2) as outlined below.

As $\theta' \alpha_g + \varepsilon_0$ is defined as the residual uncorrelated with observables, β_g comes from estimating equation (1) as a linear model in the sub sample of first time takers. Similarly, $\theta' \alpha_j + \varepsilon_{j1}$ are the residuals from the subject score equations for first time takers as there

²⁵In principle, the densities f_{θ} , f_{ε_0} , and f_{ε_t} could be non parametrically identified together with the factor loadings as in Bonhomme and Robin (2009) or Freyberger (2013).

²⁶Technical details can be found in the online appendix.

is no learning among them. Note that the correlation between error terms across the five performance equations is driven only by students' unobservables. The variance covariance matrix for these residuals is estimated by using the sample analogues. These then give a system of equations that let us estimate the factor loadings, (the α 's), the variances of $\varepsilon_0, \varepsilon_{j1} \forall t$, and the variance-covariance matrix of the θ 's.

Step 2:

Disentangling selection from learning is impossible relying only cross sectional data: we cannot compare the exam score *distributions* of first time takers to those of repeat takers and allocate the difference to learning and selection. If we had panel data these limitations would not apply.

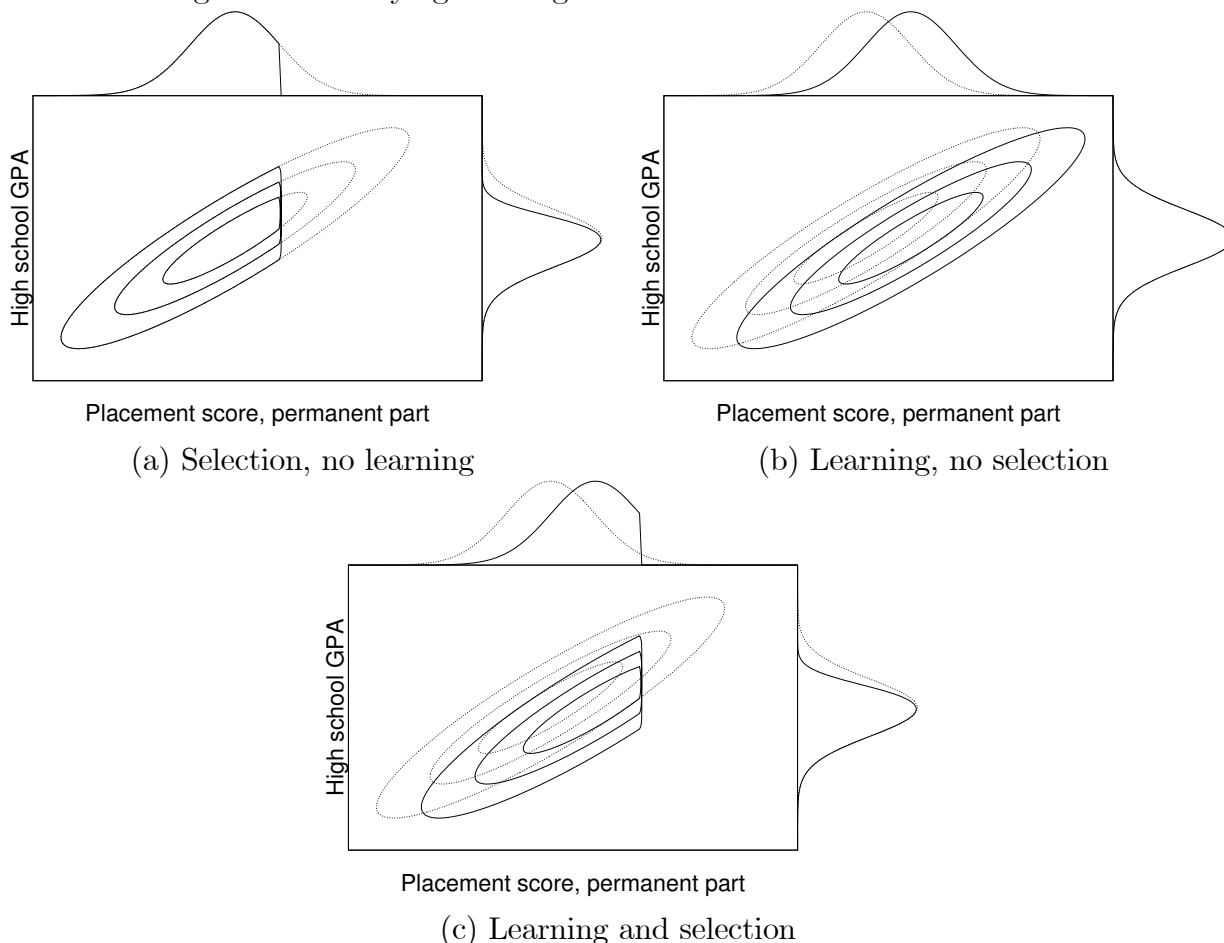
Given our data constraints, we use a novel approach that relies on the fact that GPA is not affected by retaking, which implies that the GPA distribution of repeat takers differs from that of first time takers only because of selection. The distribution of exam scores of repeat takers, in contrast, is 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 heuristically depict how selection and learning operate. In Figure 3a, we have the high school GPA and the permanent component of the placement score on the two axes for students with given observables, X . The contour curves of the joint density function are plotted for the population of first time takers as the dotted curves. The marginal distributions are depicted at the top and on the sides of the box again by dotted curves. The probability of retaking as a function of the permanent component of the placement score is given by the decision rule, i.e., $\Pr(\varepsilon_1 < e_1(\bar{s}_1, X)) = F_{\varepsilon_1}(e_1(\bar{s}_1, X))$.

Assume for illustrative purposes only, that the above decision rule takes a very special form: all those with $\bar{s}_1 \leq \bar{s}_1^*$ retake, while those above \bar{s}_1^* do not. This is shown below by truncating the joint distribution and the marginal densities as depicted by the solid curves in Figure 3a. Again, for illustrative purposes only, suppose that there is no selection and learning is positive and homogeneous for all agents who retake. Then learning just moves scores to the right as depicted by the solid lines in Figure 3b. Putting the two effects together in Figure 3c shows how both learning and selection operate under these special assumptions. Note that learning shifts the distribution of scores to the right but does not affect the high school GPA, while retaking cuts part of the distribution off.

In contrast to the examples above, we find that, in the data, the probability of retaking does not decline sharply and that learning is not homogeneous. This complicates the picture as the density functions would shift to the left while the movement to the right due to

Figure 3: Identifying learning and selection from scores and GPA



learning would be far from uniform. Nevertheless, we can use the change in the distribution of g across attempts to get a handle on the selection rule. Once we have the selection rule, we can project it on to the score distributions and obtain learning as the unaccounted part of the movement in the distribution of scores. A similar argument applies to third versus second time takers, and so on. The initiated reader will notice that what we are doing is equivalent to the first step of Hotz-Miller (1993)'s approach. In particular, we directly infer the strategies that generated the data without solving the dynamic optimization problem itself.

The retaking threshold, $e_t(\cdot)$, and the distribution of learning shocks, f_λ , are estimated via semi parametric GMM by matching the number of retakers predicted by the model to data for subsets of the GPA, placement scores, and income groups. For each income group and number of attempts, the retaking threshold is approximated by piecewise linear functions of \bar{s} on a three point grid which is specific to each group. Among the group of poor students in their second attempt for example, the three points in the grid are the 20th and

the 80th percentiles of \bar{s} within that group as well as the mean of those two scores so that the grid is regular. Similar grids are constructed for each income and number of attempts combination. Since the distributions of learning shocks are assumed to be normal, we adjust the inherited distribution of GPAs and scores according to the parameterized selection rule and parameterized distributions of the learning shocks within each relevant group.

We define cells $\Omega = \Omega_I \times \Omega_g \times \Omega_s$, where I denotes income, which can be low, middle, or high. GPAs and scores are broken down into three groups as well, leaving us with 27 cells in total. For each student who takes the exam t times, we use the retaking threshold and distribution of learning shocks hypothesized to find the probability that he ends up in cell $\Omega \in \Omega$ in attempt $t + 1$, sum this probability over all t time takers, and match this to the actual number of $(t + 1)$ takers who happen to be in Ω in the data. We do so for all cells and all attempts which gives us $(27 \times 4) = 108$ moments to match. We choose the parameters that give us the best match for these moments using unitary weights.

For example, looking at the GPA and score distributions of t^{th} time takers in the low income group, we choose the vector $\{e_t^L(\bar{s}_1), e_t^L(\bar{s}_2), e_t^L(\bar{s}_3), \mu_{\lambda t+1}, \sigma_{\lambda t+1}^2\}$ that moves the distributions so as to best match the data on the GPA and score distributions of $t + 1$ time takers. In our estimation, we assume there is no learning after attempt 4. As a result, everything is stationary after then and thus the decision rule remains the same from then on.

Step 3:

The utility function is parameterized in a flexible manner as:

$$u(s, X) = \sum_j \gamma_j \Phi\left(\frac{s - s_j}{h}\right), \gamma_j \geq 0, \sum_j \gamma_j = 1$$

The parameters (the γ 's) of the utility function are allowed to differ by income group. The normalization of $\sum_j \gamma_j = 1$ ensures that the utility at $s = \infty$ is unity. As $\Phi(\cdot)$ is increasing in s , constraining $\gamma_j \geq 0$ ensures that $u(s)$ non decreasing. The larger is h the smoother is the function. We set $h = 15$ and the number of grid points to 10.

Given a parameter vector, (γ, ψ) , and the estimate of the selection threshold obtained in step 2, we calculate the continuation values for every \bar{s} , X , and number of attempts, t which was denoted by $VC_t(\bar{s}_t, X)$. Knowing the retaking threshold $\{e_t(\bar{s}; X)\}_{t=1}^4$ and the joint distribution of shocks θ , λ and $\{\varepsilon_t\}_{t=0}^\infty$, one can find the continuation value $VC_t(\bar{s}; X, \gamma)$ for any values of γ , retaking costs ψ , discount factor δ , \bar{s} and X .²⁷ This could be done by either simulating these continuation values or by integrating over the future shocks numerically and thus obtaining them. We choose to use the latter method. Once the continuation values

²⁷This approach is based on the insights in Hotz and Miller (1993).

are known, one can estimate (γ, ψ) by matching the predicted probabilities of retaking with the actual probabilities implied by the decision rule.

Recall that the agent chooses to retake when he is better off doing so. That is

$$V_t(\bar{s}_t, \varepsilon_t, X) = \max\{u(\bar{s}_t + \varepsilon_t, X), VC_t(\bar{s}_t, X)\}.$$

Thus, the probability of retaking is the same as the probability of $u(\bar{s}_t + \varepsilon_t, X) < VC_t(\bar{s}_t, X)$. This gives us the threshold for ε_t , $\tilde{e}_t(\bar{s}_t, X)$, under which a student chooses to retake. We use this threshold $\tilde{e}_t(\bar{s}_t, X)$ in exactly the same manner as we used $e(\cdot)$ in the second stage, except that it is slightly more complicated as now $e(\cdot)$ is a function of structural parameters, not a completely flexible unknown.

For each student who takes the exam t times, we use the model to find the probability that he ends up in cell $\Omega \in \mathbf{\Omega}$ in attempt $t + 1$, sum this probability over all t time takers and match this to the actual number of $(t + 1)$ takers who happen to be in Ω in the data. We estimate the vector (γ, ψ) using the same 108 moments to match, giving them all the same weight.

The economics behind the identification of the utility is as follows: if the marginal utility of a higher score increases sharply at \bar{s} , then students close to \bar{s} will be risk loving and hence tend to want to retake the exam more than students with \bar{s} where utility of the score is less convex. Thus, the observed retaking rates pins down the curvature of the utility function. Note that the utility level cannot be interpreted as a dollar value or compared across income groups as it gives the utility of a particular score relative to the utility of the very best seat. We allow γ to vary by income and retaking costs to vary by number of attempts and income. As retaking costs are invariant across agents in a particular attempt and income group, they pin down a common effect. Differences in the probabilities of retaking across attempts help to pin down how retaking costs vary by attempt. We do not attempt to estimate δ as it is well known that it is hard to identify in such settings (see Magnac and Thesmar, 2002) and we set it at 0.9.

4 Estimation Results

4.1 Results From Step 1

In step 1, we estimate how observables are correlated with performance measures such as the GPA and the four subject exam scores in the ÖSS. We also estimate the variance of the transient shocks to these scores (ε 's), the factor loadings (α'), and the variance covariance matrix of unobserved ability (θ). We obtain these estimates for each income group and

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

Table 4: Contribution of Observables and Shocks

report them in Tables 5 and 6 below.

The key features of the estimates are the following. First, the estimates make sense overall. 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 the Math score equation. Our results indicate that the loading on the quantitative portion of ability for the science score, for example, is higher than for the social studies and vice versa for verbal ability. We also find that the variance covariance matrix for θ has more variance in q than in v . The covariance between θ_q and θ_v 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% of a standard deviation.²⁸ 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 1/3 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 4 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% 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.

²⁸This suggests that the raw correlation between income and performance often seen in the data is being captured by our other controls.

Income Outcome variable	Low				Middle				High						
	GPA	M	S	SS	T	GPA	M	S	SS	T	GPA	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).

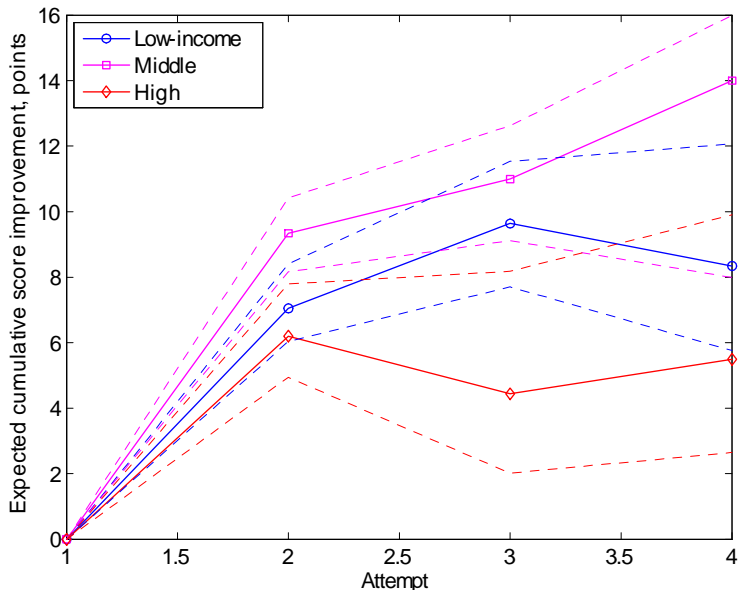
Table 5: Estimates of the GPA and the Score Equations (Continued on the Next Page)

Income	Low					Middle					High				
	GPA	M	S	SS	T	GPA	M	S	SS	T	GPA	M	S	SS	T
Outcome variable (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 6: Estimates of the GPA and the Score Equations.

Figure 4: Expected Cumulative Learning Shocks

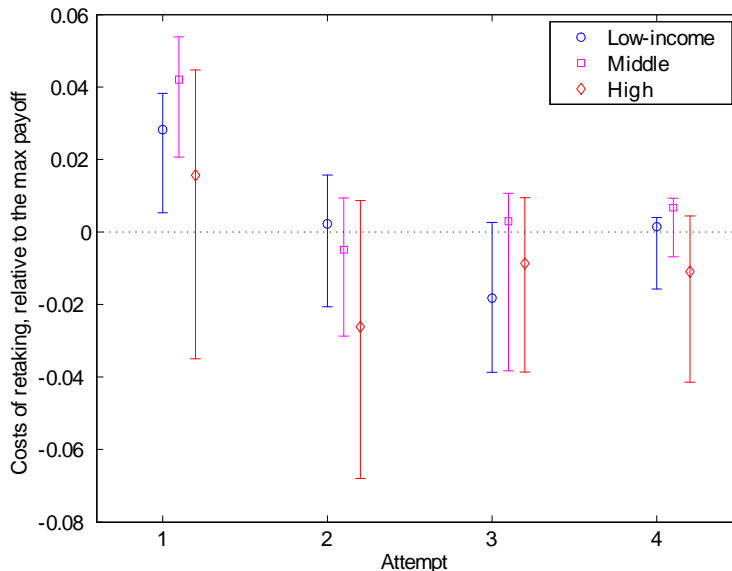


4.2 Results in Step 2

In step 2, we estimate the distribution of learning shocks. The mean learning estimates are reported in Figure 4; dashed lines depict 95% confidence bounds obtained by bootstrapping.

Students from the 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 Frisncho et. al (2013) 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 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 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, 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 standard deviation of learning shocks is about 15, roughly three times the standard deviation of the transitory shock in the placement score.



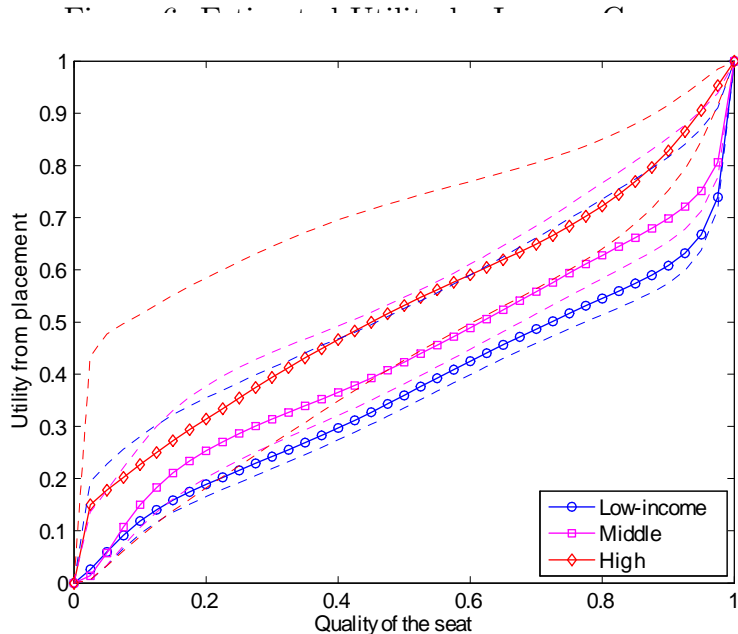
4.3 Results in Step 3

Finally, in step 3 we estimate the utility function and the retaking costs. Recall that the best placement is normalized to unity and the worst to zero. Retaking costs, depicted in Figure 5, are larger for the first retaking attempt than for later ones, where they are not statistically different from zero on average. Retaking costs in the first retaking attempt correspond to roughly 7.5% of the utility of an average placement. Surprisingly, retaking costs are not significantly different across income groups. Note that retaking costs could be monetary and/or psychic. For example, these costs could be higher on the first retake due to a stigma attached to retaking in itself.

The utility function (depicted in Figure 6) is increasing in the placement. 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 more risk loving in this region. 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.

4.4 Schooling Choices

We estimate parameters associated with schooling choices independently for the three income groups. We obtain the returns from schooling χ_{hp} from equation (7) via OLS. As 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 7 reports the estimates of schooling returns, χ_{hp} , in equation (7), 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 makes 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 1st 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 cost are obtained by using equation (8). Given our distributional assumptions, 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, and estimate the cost parameters using simulated maximum likelihood. Relative costs of various schooling options are reported in Table 7. 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

Income	Low			Middle			High		
Middle school GPA	A	B	C	A	B	C	A	B	C
Gains in score									
Anatolian (no prep)	28.9 (2.9)	27.5 (2.7)	22.9 (4.6)	27.8 (3.4)	35.5 (3.5)	34.2 (9.4)	47.4 (5.4)	30.8 (8.8)	16.4 (6.5)
Anatolian (prep)	42.6 (1.3)	43.7 (1.5)	38.7 (4.1)	44.1 (1.4)	43.7 (1.2)	40 (3.0)	45.9 (3.4)	42.5 (2.0)	38.9 (3.7)
Private, (no prep)	11 (2.0)	12.1 (2.8)	21.3 (10.0)	5.21 (2.2)	15.8 (2.3)	30.8 (11.2)	16.9 (5.4)	10.6 (3.5)	16.9 (10.7)
Private, (prep)	25.2 (1.4)	27 (1.8)	34.3 (6.5)	23.3 (1.5)	24.8 (1.3)	16.3 (6.6)	25.6 (3.5)	25.4 (2.1)	26.1 (4.1)
Public, (prep)	18 (1.5)	14.1 (1.1)	13.3 (1.8)	18.2 (1.6)	13.9 (1.0)	13.6 (2.1)	19.7 (3.7)	11.7 (1.9)	12.3 (3.5)
Costs (significant at the 1% level — in bold)									
Anatolian (no prep)	0.287 (2.07)	0.292 (3.01)	0.361 (5.50)	0.229 (0.08)	0.298 (0.17)	0.292 (0.22)	0.387 (0.18)	0.262 (1.09)	0.177 (1.92)
Anatolian (prep)	0.27 (0.49)	0.331 (1.09)	0.351 (2.63)	0.296 (0.09)	0.306 (0.02)	0.281 (0.04)	0.291 (1.77)	0.272 (1.11)	0.248 (0.65)
Private, (no prep)	0.121 (1.23)	0.197 (2.83)	0.316 (4.38)	0.050 (0.06)	0.143 (0.11)	0.267 (0.30)	0.133 (0.20)	0.088 (0.57)	0.146 (0.91)
Private, (prep)	0.159 (0.21)	0.234 (1.58)	0.382 (4.69)	0.144 (0.07)	0.173 (0.03)	0.148 (0.16)	0.138 (1.37)	0.143 (0.77)	0.151 (0.42)
Public, (prep)	0.112 (0.13)	0.085 (0.24)	0.085 (0.62)	0.115 (0.04)	0.076 (0.04)	0.075 (0.02)	0.112 (0.80)	0.039 (0.87)	0.043 (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 7: Schooling Choices: Gains and Costs.

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

5 Counterfactuals

We conduct a number of counterfactuals 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. In this part we do not incorporate effort choices. Later on, in Section 5.2, we use the extended model that allows for effort to be put in before the first attempt in the form of schooling choices and consider only a ban on retaking. Note, this is the only counterfactual we can consider that is not subject to the Lucas critique. As shown, our extended model delivers very similar results as the base one.

We compare the no-retaking scenario (labeled as 1 attempt in Table 8) and the scenario where a maximum of two attempts is allowed to the current system. We also look at the consequences of penalizing retakers by reducing their scores by 5% (labeled as 5% penalty). Finally, we experiment with doubling the weight on GPA (column x2 GPA) in the placement process. We look at the trade-off between under placement and costs of retaking. On the one hand, 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, discouraging retaking may have significant costs. On the other hand, 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 8 shows that under the 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 that retaking is excessive.

We also look at how limiting retaking affects the ability of the system to match students with seats. One way to look at mismatch is to focus at the fraction of underplaced students

	Income	Current	1 attempt	2 attempts	5% penalty	x2 GPA
Payoff	low	0.27	0.31	0.29	0.30	0.28
Payoff	medium	0.38	0.42	0.40	0.41	0.39
Payoff	high	0.55	0.59	0.57	0.59	0.56
# of attempts	low	2.28	1.00	1.40	1.52	1.97
# of attempts	medium	1.94	1.00	1.35	1.38	1.69
# of attempts	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
% underplaced	middle	35.10	16.92	25.61	22.16	28.63
% underplaced	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
% underplaced	middle	10.24	16.92	12.81	16.34	8.03
% underplaced	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 8: Policy Experiments.

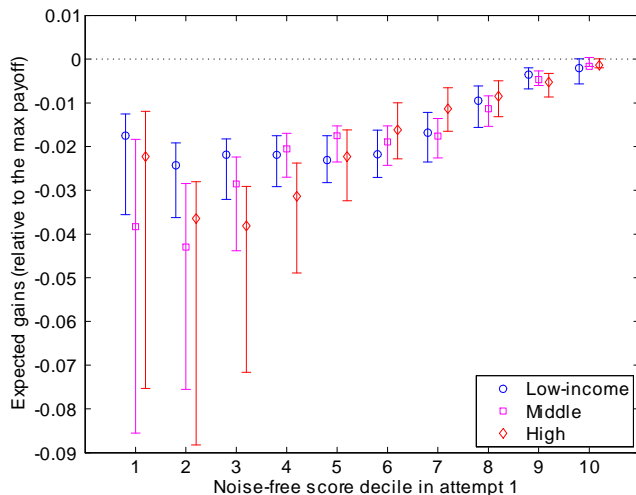
as it is the underplaced who tend to retake. We define underplacement by comparing the quality of seat to initial ability as well as ability at placement. Initial ability and ability at placement are proxied by the permanent component of the placement score in the first attempt and at the time of placement, respectively.

Rows 8-10 present the fraction of underplaced students (where underplacement is defined as being placed 5% below initial ability)²⁹. Table 8 shows that, irrespective of income, limiting retaking reduces underplacement in all the policy experiments we look at. This is counter to what intuition would suggest: in the absence of learning the underplaced would retake until they get seats comparable to their ability ranking. However, learning shocks distort the initial ranking and these distortions accumulate over time. Consequently, retaking raises mismatch relative to initial ability.

To take learning into account, we consider another definition of underplacement. In rows 11-13 we define it relative to ability at placement rather than initial ability. With this definition, limiting retaking raises underplacement as expected.

²⁹ Changing this number to 20% or 1% results in the same pattern: there is minimum mismatch in column 2, followed by column 4, followed by 3 and 5, followed by 1.

Figure 7: Gains From Banning Retaking: Partial Equilibrium



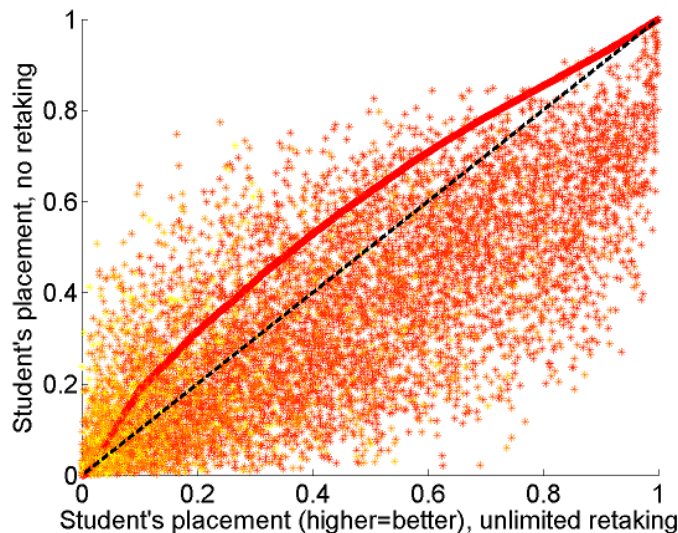
5.1 Limiting Retaking: Winners and Losers

We have shown 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.

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 Figure 7, 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. Again, this makes sense as there is excessive retaking due to the externality identified earlier.

The general equilibrium effects are illustrated in Figure 8. Each student's placement under the no-retaking rules is plotted against his placement in the current system. Retakers have a lighter color than non-retakers, with serial retakers being progressively lighter colored. It is apparent from Figure 8 that lighter colors are more prevalent towards the origin, consistent with lower-ability students retaking more often. The darker curve in the figure associates the placement of students who are placed at the first attempt under the current system with what their placement would have been in equilibrium had retaking been banned. This curve

Figure 8: Placement with and without Retaking



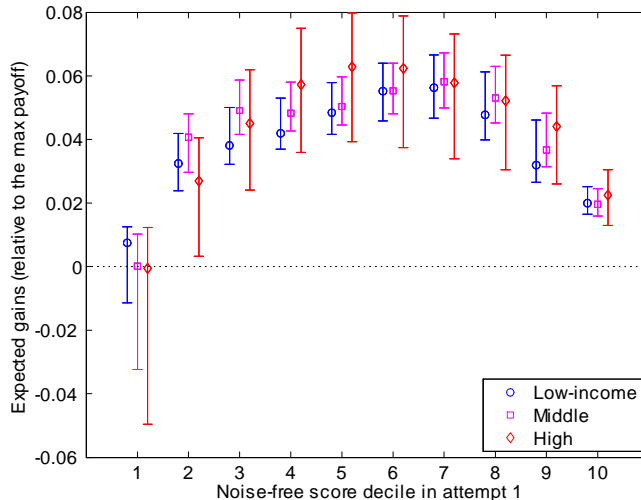
is above the 45 degree line showing that the cutoffs fall (quality of placements rise for the same score) when retaking is banned. Moreover, the fall in the cutoffs is greatest for those with mid-range placements.

In sum, the partial equilibrium consequences of preventing retaking are to reduce welfare for everyone, and more so for those with low scores. Nevertheless, the general equilibrium consequences raise welfare, and more so for those in the middle score deciles. The second effect dominates resulting in inverse U-shape gains in Figure 9. Though most agents gain ex-post, about 20% of them lose.³⁰ Some idea of this can be gleaned from Figure 8 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 9 shows that differences in gains across income groups are not significant after controlling for the initial score decile. This is somewhat unexpected as income groups do have different learning effects upon retaking as well as different retaking costs.

³⁰By ex-post we mean that we keep the shocks faced by agents constant across policy scenarios.

Figure 9: Gains from Banning Retaking: General Equilibrium, Exogenous Effort



5.2 Endogenous Effort

In the above counterfactual experiments, we assumed that the policy changes do not affect students' level of effort 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 huge increase in effort while 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 enrol into 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 less of them choose public schools. As shown in Table 9, 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 10. 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 11 depicts the point estimates and the respective 95% confidence intervals obtained via bootstrapping. The gains depicted here account for costs of effort in high school, in contrast to those plotted in Figure 9. The inverse U shaped gains are less pronounced and the average gain is halved due to welfare reducing effort effects

Income group Policy	Low		Middle		High	
	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 9: Simulated schooling choices: unlimited retaking vs 1 attempt max

Figure 10: Distribution of Placement Scores: Fixed vs Endogenous Schooling

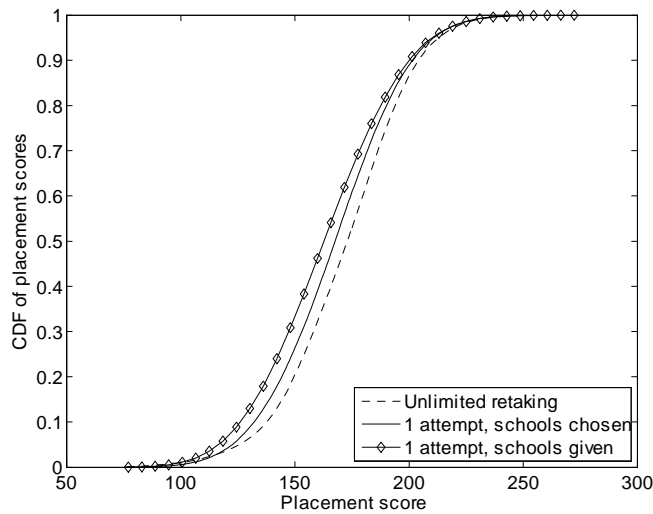
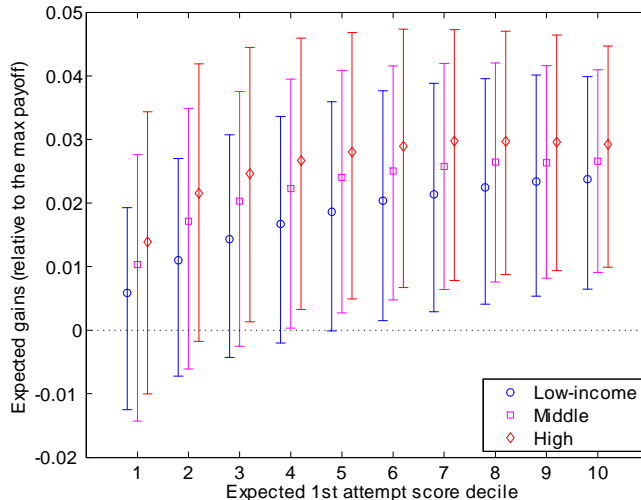


Figure 11: Gains from Banning Retaking if School Choices are Endogenous



coming from the ban. The estimates are also noisier due to the high standard errors on the cost parameters. 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.

6 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 reform in a general equilibrium setting, we can identify win-win policies like 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.³¹ 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.

³¹We should be careful in using purely vertical preferences if we were studying certain 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 modelling retaking.

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 effort effects prior to the 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 tradeoffs is a topic for future work as the computation requirements would be considerable.

References

- [1] Akerlof, George (1976) “The Economics of Caste and of the Rat Race and Other Woeful Tales”. *The Quarterly Journal of Economics*, Vol. 90, No. 4 (November).
- [2] Bonhomme, Stéphane and Jean-Marc Robin (2010). “Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics.” *Review of Economic Studies* 77 (2), 491–533.
- [3] Caner, A. and C. Okten (2010) “Risk and Career Choice: Evidence from Turkey.” *Economics of Education Review* 29 (6), 1060–1075.
- [4] Caner, A. and C. Okten (2013) “Higher education in Turkey: Subsidizing the rich or the poor?” *Economics of Education Review* 35, 75–92.
- [5] Deming, David, and Susan Dynarski (2008) "The Lengthening of Childhood." *Journal of Economic Perspectives*, 22(3): 71-92.
- [6] Fu, Qiang (2006) “A Theory of Affirmative Action in College Admissions”. *Economic Inquiry*, Vol. 44, No. 3, July, pp.420–428.
- [7] Fain, James R. (2009) “Affirmative Action Can Increase Effort”. *Journal of Labor Research*. Vol 30, pp. 168–175.
- [8] Freyberger, Joachim. (2013) “Nonparametric panel data models with interactive fixed effects” Mimeo. University of Wisconsin.
- [9] Frisanchio, Veronica, Krishna, Kala, Lychagin, Sergey and Cemile Yavas (2013) “Better Luck Next Time: Learning Through Retaking”. NBER Working Paper No. 19663.
- [10] Hatakenaka, Sachi (2006) “Higher Education in Turkey for 21st Century: Size and Composition,” November 2006, World Bank.
- [11] Hotz, V. Joseph and Robert A. Miller (1993) “Conditional Choice Probabilities and the Estimation of Dynamic Models” *The Review of Economic Studies*, Vol. 60, No. 3 (Jul., 1993), pp. 497-529.
- [12] Krishna, Lychagin and Tarasov (2015)
- [13] Magnac, Thierry, and David Thesmar (2002) “Identifying Dynamic Discrete Decision Processes”. *Econometrica*, Vol. 70, No. 2, pp. 801-816.

- [14] Mankiw, N. Gregory and Michael D. Whinston. (1986) “Free Entry and Social Inefficiency”. *The RAND Journal of Economics*, Vol. 17, No. 1 (Spring).
- [15] Saygin, P. (2011) “*Gender Differences in College Applications: Evidence from the Centralized System in Turkey.*” Working Paper.
- [16] Tansel, A. and F. Bircan (2005) “*Effect of Private Tutoring on University Entrance Examination Performance in Turkey.*” IZA Discussion Paper, No. 1609.
- [17] Tornkvist, Birgitta, and Vidar Henriksson (2004) “Repeated test taking: Differences between social groups.” EM No. 47, Umea University.
- [18] Train, Kenneth (2009). *Discrete Choice Methods with Simulation*, Cambridge University Press.
- [19] Vigdor, J. L. and C. T. Clotfelter (2003) “Retaking the SAT.” *The Journal of Human Resources* 38 (1), 1–33.

7 Appendix (For Online Publication)

7.1 Standardizing high school GPA.

The discussion below is based on that in Frisanco et. al (2013). Raw and standardized GPAs ignore potential quality heterogeneity and grade inflation across high schools. Since we are interested in obtaining a measure that will allow us to rank students on the same scale based on their high school academic performance, neither of these measures are useful. Obtaining 10/10 at a very selective school is not the same as obtaining 10/10 at a very bad school.

To deal with this issue, we constructed *school quality normalized* GPAs. Within each track k and for each school j , we define the adjustment factor, A_{jk} :

$$A_{jk} = \frac{\overline{\text{GPA}}_{jk}}{\overline{\text{Weighted Score}}_{jk}} \div \frac{\overline{\text{GPA}}_k}{\overline{\text{Weighted Score}}_k} \quad (9)$$

where $\overline{\text{GPA}}_{jk}$ and $\overline{\text{Weighted Score}}_{jk}$ are the average GPA and weighted scores for each high school and track combination. $\overline{\text{GPA}}_k$ and $\overline{\text{Weighted Score}}_k$ are the average GPA and weighted score across all comparable students from the same track.³² The numerator in (9) 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. After all, since the ÖSS is a standardized exam, $\overline{\text{Weighted Score}}_{jk}$ should be a good proxy for the true quality of the school on a unique scale. The denominator in (9) is just a constant for all the students in the same database and it takes the adjustment factor to a scale that is relative to everyone in the same track.

Define the school quality normalized GPA for student i in school j and track k as:

$$\text{GPAnorm}_{ijk} = 100 \left(\frac{\widetilde{\text{GPA}}_{ijk}}{\widetilde{\text{GPA}}_k^{\max}} \right)$$

where $\widetilde{\text{GPA}}_{ijk}$ is defined as:

$$\widetilde{\text{GPA}}_{ijk} = \left(\frac{\text{GPA}_{ijk}}{A_{jk}} \right)$$

and $\widetilde{\text{GPA}}_k^{\max}$ is just the maximum $\widetilde{\text{GPA}}_{ijk}$ in a given k . Notice that if the student is in a

³²This adjustment factor is constructed using weighted quantitative scores for Science students while Social Studies students' factor relies on weighted verbal scores. For Turkish-Math students, we use the weighted average.

school that tends to inflate the grades relative to true performance, the raw GPA of all the students in such a school will be penalized through a higher A_{jk} .

7.2 Estimating the Factor Model.

In step one we estimate the parameters of the GPA and the four subject score equations (10) using the sample of first-time takers. We obtain the estimates of β_g and β_j by running each performance measure in (10) on X .

$$\begin{aligned} g &= X'\beta_g + \theta'\alpha_g + \varepsilon_0 \\ s_{j1} &= X'\beta_j + \theta'\alpha_j + \varepsilon_{j1} \end{aligned} \tag{10}$$

Then, we use the residuals from the above regressions to pin down the factor loadings α_g and α_j , the covariance matrix of $\theta = [\theta_v, \theta_q]$ and the standard deviations of $\varepsilon_0, \varepsilon_{j1}$ where j stands for math, Turkish, social studies and science. The residuals contain the effects of unobservables and random shocks that sum up to a total of seven factors, $(\theta_v, \theta_q, \varepsilon_0, \varepsilon_{math,1}, \varepsilon_{Turk,1}, \varepsilon_{ss,1}, \varepsilon_{sc,1})$. 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 following standard assumption from the literature on factor models.

Assumption 1 The vector θ and the shocks $\varepsilon_0, \varepsilon_{math,1}, \varepsilon_{Turk,1}, \varepsilon_{ss,1}, \varepsilon_{sc,1}$ are independent of each other conditional on X .

Under the assumptions above one can identify α_g, α_j , the joint density of the common factors $[\theta_v, \theta_q]$, and the densities of the transitory shocks ε_0 and ε_{j1} non-parametrically (see Freyberger (2013) for more details). However, if we did this non-parametrically, the estimation in steps 2 and 3 would be computationally formidable. By imposing normality on the distributions of $\varepsilon_0, \varepsilon_{j1}$, and θ we circumvent this problem. As explained below, under normality all we need to estimate is the variance-covariance matrix of $\varepsilon_0, \varepsilon_{j1}$, and θ .

Let r be a vector of the five residuals from the system of equations (10)

$$r = \begin{bmatrix} \alpha_{11}\theta_v + \alpha_{12}\theta_q + \varepsilon_g \\ \alpha_{21}\theta_v + \alpha_{22}\theta_q + \varepsilon_{math,1} \\ \alpha_{31}\theta_v + \alpha_{32}\theta_q + \varepsilon_{Turk,1} \\ \alpha_{41}\theta_v + \alpha_{42}\theta_q + \varepsilon_{ss,1} \\ \alpha_{51}\theta_v + \alpha_{52}\theta_q + \varepsilon_{sc,1} \end{bmatrix}$$

Note that we can normalize these equations as follows: let $\tilde{\theta}_q = \alpha_{21}\theta_v + \alpha_{22}\theta_q$ and $\tilde{\theta}_v = \alpha_{31}\theta_v + \alpha_{32}\theta_q$. We can invert these two equations so that θ_v and θ_q are expressed in terms of $\tilde{\theta}_v$ and $\tilde{\theta}_q$. Substituting for θ_v and θ_q in terms of $\tilde{\theta}_v$ and $\tilde{\theta}_q$ into the above system gives us a normalized set of equations where the coefficients on $\tilde{\theta}_v$ and $\tilde{\theta}_q$ in the math and Turkish equations are $\alpha_{math} = [0, 1]$ and $\alpha_{Turk} = [1, 0]$ respectively.

$$r = \begin{bmatrix} \tilde{\alpha}_{11}\tilde{\theta}_v + \tilde{\alpha}_{12}\tilde{\theta}_q + \varepsilon_g \\ \tilde{\theta}_q + \varepsilon_{math,1} \\ \tilde{\theta}_v + \varepsilon_{Turk,1} \\ \tilde{\alpha}_{41}\tilde{\theta}_v + \tilde{\alpha}_{42}\tilde{\theta}_q + \varepsilon_{ss,1} \\ \tilde{\alpha}_{51}\tilde{\theta}_v + \tilde{\alpha}_{52}\tilde{\theta}_q + \varepsilon_{sc,1} \end{bmatrix}$$

Thus the normalization adds no further constraints. The covariance matrix of r can be expressed as

$$E[rr'] = \alpha' E[\theta\theta'] \alpha + I_\varepsilon$$

where I_ε is a diagonal matrix with the variances of ε 's on the diagonal. Thus, the left hand side of the above equation is a 5x5 matrix, which can be estimated from the data. 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 , and six factor loadings α . As we normalize the loadings in the math and Turkish equation to $\alpha_{math} = [0, 1]$ and $\alpha_{Turk} = [1, 0]$, four of the factor loadings in α are fixed. Therefore, we have 15 equations and 14 unknowns. We obtain the estimates of the unknown parameters using GMM with these equations as moment conditions.

7.3 Estimating the Dynamic Model of Retaking.

After estimating the factor model on the sample of first-time takers, we turn to the dynamic decision problem that models retaking behavior. We proceed in two steps. First, we use data to identify the parameters of the learning shocks and the decision rule that students follow given their ability and shocks to the score. We do so without imposing the full structure of the dynamic model. Then, using an insight from Hotz and Miller (1993), we invert the estimated decision rules to find equilibrium values of retaking and estimate the remaining parameters of the dynamic decision problem.

We start by estimating the retaking decision rules and separating learning shocks from selection into retaking. We illustrate our strategy by focusing on second-time takers; however, the same logic can be applied by induction to all subsequent retaking attempts. Knowing the actual density of placement scores and GPA, s_2 and g , among second-time takers, one can

solve an inverse problem and back out the retaking threshold e_1 and the distribution of the learning shock. Let a denote the number of attempts that a student has made. We express the joint density of placement scores and GPA among second-time takers in terms of the similar density among first-time takers. Let $\eta = \lambda_2 + \varepsilon_2$, the sum of the transitory and the permanent shock at the second attempt. Note that all densities are conditional on the covariates X though we suppress X in our notation. Let $f_{s_2, g|a=2}(s_2, g)$ denote the joint density of scores in the second attempt and GPA conditional on being a second time taker. Similarly, $f_{s_2, g, \varepsilon_1, \bar{s}_1|a=2}(s_2, g, \varepsilon_1, \bar{s}_1)$ denotes the joint density of scores on the second attempt, GPA, shocks to the score in the first attempt and the permanent component of the first time score conditional on being a second time taker. Then,

$$\begin{aligned}
f_{s_2, g|a=2}(s_2, g) &= \iint f_{s_2, g, \varepsilon_1, \bar{s}_1|a=2}(s_2, g, \varepsilon_1, \bar{s}_1) d\varepsilon_1 d\bar{s}_1 \\
&= \iint f_{\eta, g, \varepsilon_1, \bar{s}_1|a=2}(s_2 - \bar{s}_1, g, \varepsilon_1, \bar{s}_1) \left| \frac{\partial(\eta, g, \varepsilon_1, \bar{s}_1)}{\partial(s_2, g, \varepsilon_1, \bar{s}_1)} \right| d\varepsilon_1 d\bar{s}_1 \\
&= \iint f_{g, \varepsilon_1, \bar{s}_1|a=2}(g, \varepsilon_1, \bar{s}_1) f_\eta(s_2 - \bar{s}_1) d\varepsilon_1 d\bar{s}_1 \\
&= \iint f_{g, \varepsilon_1, \bar{s}_1|a=1}(g, \varepsilon_1, \bar{s}_1) I(\varepsilon_1 < e_1(\bar{s}_1)) f_\eta(s_2 - \bar{s}_1) d\varepsilon_1 d\bar{s}_1 \frac{1}{\Pr\{a = 2\}} \quad (11)
\end{aligned}$$

In the first line we integrate the joint distribution of second-time score, s_2 , GPA, g , shocks ε_1 and first-time noise-free score \bar{s}_1 among second-time takers over the last two variables. Then, in the second line, we express the density $f_{s_2, g, \varepsilon_1, \bar{s}_1|a=2}$ via the density $f_{\eta, g, \varepsilon_1, \bar{s}_1|a=2}$ using the fact that $\eta = s_2 - \bar{s}_1$. In the third line we use independence of η to separate the marginal density of η from the joint density of \bar{s}_1, g and ε_1 . This requires the following assumption:

Assumption 2 The distribution of learning shocks, λ_t , and idiosyncratic shocks, ε_t , are independent of the history conditional on observables, X .

Finally, in the last line we go from conditioning on being a second-time taker to being first-time taker. The density of second-time takers is merely the density of first-time takers who meet the selection rule scaled by the fraction of first-time takers who retake (in other words, we use Bayes rule).

Note that the estimate of $f_{g, \varepsilon_1, \bar{s}_1|a=1}$ can be obtained from the factor model; the probability $\Pr\{a = 2\}$ is the retaking rate directly observable in the data. Given a decision rule e_1 and a distribution of the second attempt shocks to the placement score, f_η , one can predict the distribution of scores and GPA among the second-time takers. The estimates of e_1 and f_η are obtained by fitting this prediction to the data. For each of the income groups, we

partition the set of second-time takers by their GPA using GPA terciles as cutoffs. Each of the resulting three subsets is further cut into three smaller sets of equal sizes using placement score terciles. We use equation 11 to predict the numbers of retakers in the nine subsets of the score-GPA space defined above and match them to the numbers of retakers in the data. This gives us nine moment equations per income group, which we use to obtain GMM estimates of e_1 (3 unknowns) and the parameters of f_η (2 unknowns).

Note that one needs additional assumptions to separate learning, λ_2 , from idiosyncratic shocks, ε_2 , given the density of their sum, f_η . Assuming that ε_2 is drawn from the same distribution as ε_1 (identified from the factor model) and that ε_2 is orthogonal to λ_2 , one can identify the distribution of λ_2 by deconvolution.

Assumption 3 The distribution of idiosyncratic shocks ε_t does not vary across attempts.

Learning shocks λ_t are independent of ε_t conditional on X .

The above argument can easily be generalized to attempts 3 and 4. It has to be appropriately modified for attempts 5 onwards as agents who take the exam five times or more are pooled in the data (that is, we cannot separate 5-time takers and 6-time takers). Let $f_{\bar{s},g|a>4}(\bar{s}, g)$ be the density of the permanent component of the placement score and GPA in the population of 5 and more time takers. For simplicity, we assume that there is no learning beyond attempt 4. As learning is the only reason why \bar{s} evolves, each student's \bar{s} is fixed after the fourth attempt. Moreover, the decision rule is unchanging after the 4th attempt, as the student faces a stationary environment in the absence of learning shocks. The aggregate group of 5+ time takers is composed of those 4 and 5+-time takers who decided to retake in the past. Thus,

$$f_{\bar{s},g|a>4}(\bar{s}, g) = (f_{\bar{s},g|a=4}(\bar{s}, g) + f_{\bar{s},g|a>4}(\bar{s}, g)) F_{\varepsilon_1}(e_4(\bar{s})).$$

Rearranging gives:

$$f_{\bar{s},g|a>4}(\bar{s}, g) = \frac{f_{\bar{s},g|a=4}(\bar{s}, g) F_{\varepsilon_1}(e_4(\bar{s}))}{1 - F_{\varepsilon_1}(e_4(\bar{s}))}. \quad (12)$$

This equation can be used to obtain the distribution of scores and grades among students who retake 5 times and more. Note that $f_{\varepsilon_t}(\varepsilon_t)$ is the same for all t by assumption. For this reason we drop the subscript t below.

$$\begin{aligned}
f_{s,g|a>4}(s, g) &= \int f_{\bar{s},g|a>4}(s - \varepsilon, g) f_\varepsilon(\varepsilon) d\varepsilon \\
&= \int f_{\bar{s},g|a>4}(\bar{s}, g) f_\varepsilon(s - \bar{s}) d\bar{s} \\
&= \int \frac{F_\varepsilon(e_4(\bar{s}))}{1 - F_\varepsilon(e_4(\bar{s}))} f_{\bar{s},g|a=4}(\bar{s}, g) f_\varepsilon(s - \bar{s}) d\bar{s}
\end{aligned}$$

In the first line above we get the distribution of $s = \bar{s} + \varepsilon$ from that of \bar{s} and ε using independence. In the third line we substitute using equation (12). The density $f_{\bar{s},g|a=4}$ is identified as outlined above. Knowledge of F_ε and the selection cutoff e_4 gives us the right-hand side of the above equation. We identify e_4 by matching the right-hand side to the empirical distribution of scores and GPAs of 5 and more time takers.

After obtaining the equilibrium decision rules we are in a position to estimate the dynamic model's parameters by following along the same lines as Hotz and Miller (1993) in their second step. First, we use the decision rules to compute continuation values in the dynamic model of retaking. Then, assuming that each student maximizes his/her expected welfare, we pin down the retaking costs and the placement utility function that best explain the retaking rates in the nine subsets of placement scores and GPA defined above. We implement this step in the GMM framework and use bootstrapping to obtain standard errors.

7.4 Simulation Algorithm.

In our counterfactual policy experiments we simulate retaking behavior in the steady states that arise after policy changes. Given the model's parameters, a steady state can be characterized by a distribution of placement scores *of those students who choose to be placed* which pins down the allocation of students (identified by their score) to schools. This distribution is a general equilibrium object; it changes whenever the policy environment is being changed. Among other things, the distribution of scores captures the level of competition in the exam. Given this distribution, each student knows what seat his score will buy, should he decide to be placed.

The objective of the simulation algorithm is to find the above conditional distribution. The c.d.f. of this distribution is approximated on a uniform grid of 36 points. The numerical procedure is as follows:

1. Given a candidate c.d.f., the algorithm solves each student's dynamic decision problem
2. The solution to the dynamic problem is used to simulate student scores and retaking

decisions

3. Scores of retakers are used to compute the new conditional c.d.f.
4. If the candidate c.d.f. and the new one are close, the algorithm stops. Otherwise, a new candidate c.d.f. is tried.