Retaking in High Stakes Exams: Is Less More?*

Kala Krishna, Sergey Lychagin, Veronica Frisancho

19th September 2014

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 much of the world is also similarly structured and is fiercely

---

*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 Central European University, the 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 authors declare that they have no relevant material or financial interests that relate to the research described in this paper.
competitive. Students often 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.\footnote{See the article entitled “Trading Delayed as 650,000 South Koreans Take College Test”. Bloomberg News, November 6, 2013. Also see the article “South Korea’s dreaded college entrance exam is the stuff of high school nightmares, but is it producing "robots"?”. CBS News, November 7, 2013.} 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.\footnote{http://www.huffingtonpost.com/2012/07/02/china-test_n_1644306.html} 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.\footnote{Due 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.} 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. A survey by Bound et al (2009) documents a rise in applicant activities aimed at impressing college
admissions officials. Our results suggest that an unrestricted opportunity to retake admission tests could have adverse consequences in such circumstances. Retaking is also related to the issue of “red-shirting”\footnote{This refers to holding an athlete back till he is stronger and more able to compete.}. 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 is no systematic analysis of its costs and benefits in centralized placement systems. 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? When would gains tend to exceed the losses? 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 4\textsuperscript{th} 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\footnote{See Hatakenaka (2006) for more on the challenges for the Turkish higher education system.}. 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. Only when selection is controlled for the true learning gains can be pinned down.

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 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, given cutoff scores, he obtains his placement faster without retaking. At the same time, competition for placement is less fierce and the same score is worth more when retaking is restricted. This occurs both because fewer students compete for placement at any time, and because there is less studying to the exam between attempts. In our simulations, this general equilibrium effect dominates, so that everyone is in fact better off by restricting retaking. Nevertheless, if students are naive and cannot anticipate general equilibrium effects of such reforms, they will resist the restrictions.

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 discriminate in favor of more advantaged groups as these have lower costs of retaking, retake more often, and so get higher maximum scores across attempts. However, as they do not model the equilibrium in their paper, they are forced to assume that schools do not change their admission rules in their counterfactuals. However, we show that these general equilibrium effects are critical; had we made the same assumption

\[\text{This is another example of the fallacy of composition at work.}\]
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. Our paper models the contest as an anonymous game where effort is not per se desirable and students take cut-offs to be admitted as given. This is analogous to monopolistic competition where firms take the price index as given.

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.

Footnotes:

7 For instance, see Fu (2007) and Fain (2009).
8 In fact, retaking being excessive in our model is analogous to the result of Mankiw and Whinston (1986) on excessive entry with homogeneous firms and monopolistic competition. Just 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. 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

\(^9\)For more on how these scores differ from each other see Frisancho et. al (2013).
\(^{10}\)Only those students who obtain more than a certain score are eligible to submit preferences.
enough to hurt their placement a lot. In the data, the placement rate of those who retake when enrolled is 15 percent while that of students who are not enrolled is 33 percent. 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.\footnote{Had we not assumed placement was terminal, we would have complicated the model a lot.}

Our data covers a random sample of about 42,731 students who took the ÖSS in 2002 and who were in the Science track. ÖSYM data comes from three sources: students’ application forms, a survey given in 2002, and administrative data on high school GPA and scores in each part of the exam. After cleaning the data, dealing with some minor inconsistencies (4%) 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).\footnote{It 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.}

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

Table 1: Summary Statistics
In 2002, more than 50% of the science track applicants were repeat takers. According to our data, approximately 80% of retakers are not employed at the time of the exam. This suggests that a large fraction of the young population stays out of the labor force or the formal education system while retaking.

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, say because youth unemployment is high or because there is no stigma attached to retaking, 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. 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 fraction of students in low, middle and high income groups for each retaking attempt. 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

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

14 Our definition of income groups splits the population into three roughly equal parts. Students in the low income group report monthly household income of less than 250 Turkish lira (TL). Households earning more than 500 TL are classified as high-income ones. Those in between 250 TL and 500 TL 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.
restricting retaking may affect the poor more adversely than other groups.

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

Table 2: Retaking Rates by Income

1.2 Learning and Selection into Retaking

As we mention above, retaking rates may vary across students for different reasons. One reason could be that socioeconomic groups vary in their access to tutoring, returns from higher education and expected gains from learning between retaking attempts. Another explanation is that retaking rates are determined by the students’ initial abilities, which are formed in school or in early childhood.

In Section 3 we show how students’ ability affect retaking rates and how much scores improve between attempts using the distributions oh high school GPA and exam scores. Before we do this, here 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 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).

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. It is important to not only know the size of the parameters, but how they vary across
income groups. This is why we need to estimate the model structurally and pin down the costs and benefits of retaking separately for each group. This is what we turn to next.

2 Modelling the Turkish System

We make the following key assumptions: i) students know their own ability though this is unobserved by the econometrician, ii) learning is a draw from a distribution that is allowed to vary with observables, and iii) performance in high school and on 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.

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, these GPAs are normalized to be comparable using the school’s performance in the university entrance exam to do so. Details on how this is done are in the Appendix. We will
postulate that the normalized high school GPA for a given student\textsuperscript{15} is given by

\[ g = X' \beta_g + \theta' \alpha_g + \varepsilon_0. \]  

(1)

where \( X \) is a vector of individual characteristics 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 innate ability, \( \theta_q \) and \( \theta_v \), 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. \( \theta \) is observed by the student, but unobserved by the econometrician. The shock, \( \varepsilon_0 \), captures the randomness associated with the GPA. The distributions of \( \theta \) and \( \varepsilon_0 \) 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

\[ s_{jt} = X' \beta_j + \theta' \alpha_j + \sum_{\tau=2}^{t} \lambda_{jt} + \varepsilon_{jt} \]  

(2)

where \( \varepsilon_{jt} \) is the corresponding error term for the student’s score in subject \( j \) (social studies, science, Turkish and math) and attempt \( t \).\textsuperscript{16} We assume that the \( \varepsilon_{jt} \)’s are i.i.d. conditional on \( X \). \( \lambda_{jt} \) 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 \( \theta \)’s, \( \lambda \)’s and \( \varepsilon \)’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, \( \alpha_g \) and \( \alpha_j \), do not vary across students in a group or attempts. They do 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.\textsuperscript{17}

\textsuperscript{15}We dispense with individual subscripts for ease of notation.

\textsuperscript{16}Note that \( \theta \) and \( \varepsilon_{jt} \) differ in that the draw of \( \theta \) is the same for a given student for the GPA as well as exam scores, while the draws of the \( \varepsilon_0 \) and \( \varepsilon_{jt} \)’s varies across them.

\textsuperscript{17}This normalization is without loss of generality. See the Appendix for details.
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.

$$
s_t = w_gg + \sum_j w_j s_{jt}
$$

$$
= w_gg + X'\sum_j w_j \beta_j + \theta' \sum_j w_j \alpha_j + \sum_j \sum_{\tau=2}^t w_j \lambda_{j\tau} + \sum_j w_j \varepsilon_{jt}
$$

$$
= w_gg + X'\beta + \theta' \alpha + \sum_{\tau=2}^t \lambda_{\tau} + \varepsilon_t
$$

(3)

where the aggregate transitory shock to the $t^{th}$ placement score is denoted as $\varepsilon_t$ and $\lambda_{\tau}$ captures the (permanent) learning shocks in the $t^{th}$ attempt.\footnote{In what follows, we assume that the variance of $\varepsilon_t$ is unaffected by $t$. However, note that the variance of $\varepsilon_t$ and $\varepsilon_0$ can differ. This makes sense: the GPA is accumulated over the year so it may have a smaller transitory shock than the exam score.}\footnote{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 $X$, $\theta$, and $\varepsilon_0$. 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 in his first attempt, the student learns his $\varepsilon_1$. Students put down their preferences only after knowing their score in the exam. Knowing his $X$, $\theta$, $\varepsilon_0$ and $\varepsilon_1$, the student decides whether or not to retake the exam. If the student does not retake, he accepts his most preferred outcome, which includes 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 in upon deciding to retake. This captures the fact that the exam is given only once a year. Second time takers study for the exam and so learn their $\lambda_2$. Upon learning their score in the second attempt, they observe their $\varepsilon_2$. A similar time line occurs for later attempts. The future is discounted at a common rate of $\delta$ by all students.
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 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 their scores, \( G(s) \), one can derive \( u(s, X) \) as\(^{22}\):

\[
    u(s, X) = U(G(s); X) .
\]

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

\(^{19}\)Note that even in more general settings were preferences are not strictly vertical, the indirect utility is increasing in \( r \) as a higher rank allows for more options.

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

\(^{21}\)For example, is students retake based on both preference and performance shocks, retaking choices would be much harder to estimate.

\(^{22}\)The mass of students who do not retake is normalized to one in the steady state.
0) and unity \((U(1, X) = 1)\), respectively.

The distribution of scores \(G(s)\) is an equilibrium outcome. A change in the rules of 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 you 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 will be entirely determined by the permanent component of the score, \(\bar{s}_t\), and the corresponding transitory shock, \(\varepsilon_t\). 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 \(V C_t(\bar{s}_t, X)\) be the continuation payoff. As usual:

\[
V_t(\bar{s}_t, \varepsilon_t, X) = \max \{u(\bar{s}_t + \varepsilon_t, X), V C_t(\bar{s}_t, X)\} \tag{5}
\]

where

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

where \(\psi_t(X)\) denotes that retaking costs may vary with observables (such as income group as in our main econometric specification) and \(E_{\varepsilon_{t+1}, \lambda_{t+1}|X}\) emphasizes that the distributions of \(\varepsilon\) and \(\lambda\) 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 \(\varepsilon_t\) is below a threshold. Thus, a student’s decision follows a simple rule: retake after the \(t^{th}\) attempt if

\[
\varepsilon_t < e_t(\bar{s}_t, X) \tag{6}
\]

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,
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 $u(s, X)$. We normalize the number of students placed in steady state to unity.

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 (thereby taking up their outside option). 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 addition, the path is generated by their behavior. Note that $U(r, X)$ is the primitive utility, while $u(s, X)$ is an equilibrium outcome.

### 2.3 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) is drawn from a normal distribution, $N[130, 25]$, with mean and 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, $U(r) = r^\xi$, and $\xi = 2$. 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 $Var[\varepsilon] = 25$.

In columns 3 and 4 of Table 3 we can see that when the variance of $\varepsilon$ or $\lambda$ rises, 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 an attempt more likely to retake. However, the negative externality retakers inflict on others makes
expected utility fall when retaking rises. In column 5 we raise retaking costs and this reduces retaking while raising expected utility. This suggests that polices that reduce retaking costs, such as more frequent exams, may be a bad idea. In column 6, we make agents risk averse. 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. Thus, the seemingly reasonable response of increasing seats as a response to a backlog of students might actually make things worse.

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, 2014) 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 more attractive. In this case, banning retaking results in the three highest score deciles gaining, but the majority loose 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 redistributational aspects of such a reform: the majority may in fact prefer unlimited retaking. In the last example, however, the losses of the majority are less than the gains of the minority.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>$Var[\varepsilon] \times 4$</th>
<th>$Var[\lambda] \times 4$</th>
<th>$\psi \times 2$</th>
<th>$U(r) = \sqrt{r}$</th>
<th>+10% top seats</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$ (attempts)</td>
<td>1.19</td>
<td>1.48</td>
<td>1.57</td>
<td>1.03</td>
<td>1.03</td>
<td>1.21</td>
</tr>
<tr>
<td>$E$ (utility)</td>
<td>0.32</td>
<td>0.29</td>
<td>0.29</td>
<td>0.33</td>
<td>0.66</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 3: Simulated Comparative Statics
If, in addition to heterogeneity in agents and schools, there are gains from matching better agents to better schools, retaking may help improve the match. In this environment, banning retaking can reduce aggregate welfare.

Figure 2: Preventing Retaking: Welfare Consequences

![Graphs showing welfare consequences for different attempts and risk preferences.]

(a) Baseline Case  (b) Risk Loving Agents

3 Identification

Our goal is to estimate the key structural parameters of the model by income group. The estimation strategy we follow has 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 $f_{\epsilon_0}$, $f_{\epsilon_t}$, and $f_{\theta}$, 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 $f_{\epsilon_t}$, $f_{\epsilon_0}$, and $f_{\epsilon_t}$ are normal which makes their estimation straightforward.

In step 2, we estimate the selection cutoffs defined in equation (6) and the distributions of the learning shocks, $\lambda_t$. Step 3 deals with the dynamic component; in this step, we estimate the costs of retaking, $\psi$, and the utility function defined in equation (4). We discuss the intuition behind each step below.

**Step 1:**

23In principle, the densities $f_{\theta}$, $f_{\epsilon_0}$, and $f_{\epsilon_t}$ could be non parametrically identified together with the factor loadings as in Bonhomme and Robin (2009) or Freyberger (2013).
In Turkey, practically every high school senior takes the university entrance exam. For this reason, the sub-sample of first time takers is free of selection. By imposing normality on the distributions of $\varepsilon_0, \varepsilon_t \forall t$, and $\theta$, we can easily estimate the distributions of $\varepsilon_0, \varepsilon_t \forall t$, and $\theta$ 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, $\beta_g$ 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 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 $\alpha$’s), the variances of $\varepsilon_0, \varepsilon_{j1} \forall t$, and the variance covariance matrix of the $\theta$’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 constraints would not apply.

Given our data constraints, we rely 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 sample 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.,
\[ \Pr(\varepsilon_1 < e_1(\tilde{s}_1, X)) = F_{\varepsilon_1}(e_1(\tilde{s}_1, X)). \]

Figure 3: Identifying learning and selection from scores and GPA

(a) Selection, no learning
(b) Learning, no selection
(c) Learning and selection

Assume for illustrative purposes only, that the selection rule takes a very special form: all those with \( \tilde{s}_1 \leq \tilde{s}_1^* \) retake, while those above \( \tilde{s}_1^* \) 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, better students retake less often 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 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 onto 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, $c_t(\cdot)$, and the distribution of learning shocks, $f_{\lambda}$, 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 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 $\Omega$ 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 $\{e_t^L(\bar{s}_1), e_t^L(\bar{s}_2), e_t^L(\bar{s}_3), \mu_{M+1}, \sigma_{M+1}^2\}$ 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 $\gamma$'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 $\Phi(.)$ 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, ($\gamma, \psi$), 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 $VC_t(\bar{s}, X)$. Knowing the retaking threshold $\{c_t(\bar{s}; X)\}_{t=1}^4$ and the joint distribution of shocks $\theta$, $\lambda$ and $\{\epsilon_t\}_{t=0}^\infty$, one can find the continuation value $VC_t(\bar{s}; X, \gamma)$ for any values of $\gamma$, retaking costs $\psi$, discount factor $\delta$, $\bar{s}$ and $X$. 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 are known, one can estimate ($\gamma$, $\psi$) 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

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

Thus, the probability of retaking is the same as the probability of $u(\bar{s}_t + \epsilon_t, X) < VC_t(\bar{s}_t, X)$. This gives us the threshold for $\epsilon_t$, $\tilde{c}_t(\bar{s}_t, X)$, under which a student chooses to retake. We use this threshold $\tilde{c}_t(\bar{s}_t, X)$ in exactly the same manner as we used $c(.)$ in the second stage, except that it is slightly more complicated as now $c(.)$ 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 $\Omega \in \Omega$ in attempt $t + 1$, sum this probability over all

\[^{24}\text{This approach is based on the insights in Hotz and Miller (1993).}\]
time takers and match this to the actual number of \((t + 1)\) takers who happen to be in \(\Omega\) in the data. We estimate the vector \((\gamma, \psi)\) 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 \(\gamma\) 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 \(\delta\) 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 \((\varepsilon)'s\), the factor loadings \((\alpha)'s\), and the variance covariance matrix of unobserved ability \((\theta)\). We obtain theses estimates for each income group and report them in Tables 4 and 5 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 \(\theta\) has more variance in \(q\) than in \(v\).
<table>
<thead>
<tr>
<th>Income Outcome variable</th>
<th>GPA</th>
<th>M</th>
<th>S</th>
<th>SS</th>
<th>T</th>
<th>GPA</th>
<th>M</th>
<th>S</th>
<th>SS</th>
<th>T</th>
<th>GPA</th>
<th>M</th>
<th>S</th>
<th>SS</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father's occupation (base category – employer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Works for wages/salary</td>
<td>-0.55</td>
<td>0.10</td>
<td>0.51</td>
<td>1.08</td>
<td>0.25</td>
<td>1.11**</td>
<td>2.02**</td>
<td>2.14**</td>
<td>0.92</td>
<td>-0.48</td>
<td>0.80**</td>
<td>1.20**</td>
<td>1.65**</td>
<td>1.74**</td>
<td>1.09**</td>
</tr>
<tr>
<td>Self-employed</td>
<td>-0.58</td>
<td>0.30</td>
<td>0.50</td>
<td>1.06</td>
<td>0.35</td>
<td>1.08**</td>
<td>2.20**</td>
<td>1.93**</td>
<td>0.10</td>
<td>-0.97</td>
<td>0.29</td>
<td>0.23</td>
<td>1.08**</td>
<td>1.64**</td>
<td>1.42**</td>
</tr>
<tr>
<td>Unemployed/not in LF</td>
<td>-0.74</td>
<td>0.16</td>
<td>0.52</td>
<td>1.16</td>
<td>0.14</td>
<td>0.87</td>
<td>1.96*</td>
<td>1.94*</td>
<td>0.55</td>
<td>-0.87</td>
<td>0.20</td>
<td>-0.25</td>
<td>0.78</td>
<td>2.90**</td>
<td>0.91</td>
</tr>
<tr>
<td>Mother's occupation (base category – employer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Works for wages/salary</td>
<td>3.88**</td>
<td>0.73</td>
<td>2.53</td>
<td>1.53</td>
<td>-0.22</td>
<td>-1.32</td>
<td>-4.38</td>
<td>-2.59</td>
<td>-4.05</td>
<td>1.09</td>
<td>1.46**</td>
<td>2.31</td>
<td>1.94</td>
<td>1.92</td>
<td>2.33</td>
</tr>
<tr>
<td>Self-employed</td>
<td>2.47**</td>
<td>-1.19</td>
<td>0.57</td>
<td>0.24</td>
<td>-2.74</td>
<td>-1.17</td>
<td>-4.06</td>
<td>-1.64</td>
<td>-3.02</td>
<td>0.54</td>
<td>1.96**</td>
<td>3.81**</td>
<td>2.84*</td>
<td>1.06</td>
<td>1.40</td>
</tr>
<tr>
<td>Unemployed/not in LF</td>
<td>3.11**</td>
<td>0.26</td>
<td>1.55</td>
<td>0.56</td>
<td>-1.57</td>
<td>-1.02</td>
<td>-3.66</td>
<td>-1.81</td>
<td>-3.07</td>
<td>1.56</td>
<td>1.92**</td>
<td>2.72*</td>
<td>1.94</td>
<td>1.89</td>
<td>3.00**</td>
</tr>
<tr>
<td>Father's education (base category primary school or lower level)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle/High school</td>
<td>-0.11</td>
<td>0.09</td>
<td>0.02</td>
<td>0.49</td>
<td>0.23</td>
<td>0.27</td>
<td>0.63</td>
<td>0.41</td>
<td>0.13</td>
<td>-0.21</td>
<td>-0.45</td>
<td>-1.06</td>
<td>-0.11</td>
<td>-0.67</td>
<td>0.39</td>
</tr>
<tr>
<td>2-year higher education</td>
<td>-0.15</td>
<td>0.25</td>
<td>1.11</td>
<td>0.82</td>
<td>1.65</td>
<td>0.69**</td>
<td>1.65**</td>
<td>1.21*</td>
<td>0.15</td>
<td>0.43</td>
<td>0.25</td>
<td>0.10</td>
<td>1.21</td>
<td>0.06</td>
<td>1.42**</td>
</tr>
<tr>
<td>College/Master/Phd</td>
<td>0.88</td>
<td>1.85</td>
<td>2.06*</td>
<td>2.83*</td>
<td>2.34*</td>
<td>0.98**</td>
<td>2.04**</td>
<td>2.17**</td>
<td>1.24**</td>
<td>1.46**</td>
<td>0.61**</td>
<td>0.99</td>
<td>2.04**</td>
<td>1.22</td>
<td>2.18**</td>
</tr>
<tr>
<td>Missing</td>
<td>-0.37</td>
<td>-0.72</td>
<td>-0.09</td>
<td>-1.32*</td>
<td>-1.36</td>
<td>0.26</td>
<td>0.79</td>
<td>0.56</td>
<td>0.54</td>
<td>-0.83</td>
<td>-0.60</td>
<td>-2.74*</td>
<td>-1.38</td>
<td>-0.65</td>
<td>-0.08</td>
</tr>
<tr>
<td>Mother's education (base category primary school or lower level)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle/High school</td>
<td>-0.37</td>
<td>-0.12</td>
<td>-0.47</td>
<td>-0.15</td>
<td>0.82</td>
<td>-0.36</td>
<td>-0.88**</td>
<td>-0.51</td>
<td>-0.38</td>
<td>0.32</td>
<td>-0.44**</td>
<td>-0.56</td>
<td>-0.99**</td>
<td>0.37</td>
<td>-0.21</td>
</tr>
<tr>
<td>2-year higher education</td>
<td>-1.32</td>
<td>0.26</td>
<td>-3.54</td>
<td>0.27</td>
<td>3.36</td>
<td>0.13</td>
<td>-0.14</td>
<td>0.50</td>
<td>0.67</td>
<td>2.18**</td>
<td>0.60</td>
<td>0.88</td>
<td>0.50</td>
<td>0.98</td>
<td>1.70**</td>
</tr>
<tr>
<td>College/Master/Phd</td>
<td>-2.08</td>
<td>-3.87</td>
<td>-1.00</td>
<td>-0.79</td>
<td>0.21</td>
<td>-0.14</td>
<td>-1.31</td>
<td>-0.49</td>
<td>0.40</td>
<td>0.40</td>
<td>0.81**</td>
<td>1.03</td>
<td>1.13</td>
<td>2.05**</td>
<td>1.94**</td>
</tr>
<tr>
<td>Missing</td>
<td>0.53</td>
<td>0.68</td>
<td>-0.38</td>
<td>1.17</td>
<td>-1.68*</td>
<td>-0.47</td>
<td>-0.94</td>
<td>-0.43</td>
<td>-1.33</td>
<td>1.06</td>
<td>0.68</td>
<td>2.55</td>
<td>1.66</td>
<td>0.91</td>
<td>1.01</td>
</tr>
<tr>
<td>Likely source of income in college (base category family, rental income or scholarship)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td>-0.55**</td>
<td>-1.34**</td>
<td>-1.25**</td>
<td>-0.64</td>
<td>-0.96**</td>
<td>-0.47**</td>
<td>-0.43</td>
<td>-0.96**</td>
<td>-0.78*</td>
<td>-0.88**</td>
<td>-0.69**</td>
<td>-1.06**</td>
<td>-0.80**</td>
<td>0.02</td>
<td>-0.30</td>
</tr>
<tr>
<td>Loan</td>
<td>-0.03</td>
<td>0.36</td>
<td>0.20</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.18</td>
<td>0.69*</td>
<td>0.32</td>
<td>-0.77**</td>
<td>-0.51</td>
<td>0.11</td>
<td>0.78*</td>
<td>0.39</td>
<td>-0.70</td>
<td>-0.68*</td>
</tr>
<tr>
<td>Other</td>
<td>-0.60</td>
<td>-0.51</td>
<td>-0.39</td>
<td>0.08</td>
<td>-0.54</td>
<td>-0.75*</td>
<td>-0.64</td>
<td>-1.38**</td>
<td>-0.67</td>
<td>-0.59</td>
<td>-0.77*</td>
<td>-0.13</td>
<td>-1.58*</td>
<td>-1.83*</td>
<td>-0.79</td>
</tr>
</tbody>
</table>

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 4: Estimates of the GPA and the Score Equations (Continued on the Next Page)
<table>
<thead>
<tr>
<th>Income</th>
<th>Low GPA</th>
<th>Low M</th>
<th>Low S</th>
<th>Low SS</th>
<th>Low T</th>
<th>Med GPA</th>
<th>Med M</th>
<th>Med S</th>
<th>Med SS</th>
<th>Med T</th>
<th>High GPA</th>
<th>High M</th>
<th>High S</th>
<th>High SS</th>
<th>High T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet access (base category – at home)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes, not at home</td>
<td>0.56</td>
<td>1.60</td>
<td>0.49</td>
<td>-0.26</td>
<td>0.38</td>
<td>0.06</td>
<td>-0.02</td>
<td>0.49</td>
<td>-0.63</td>
<td>-0.33</td>
<td>-0.40</td>
<td>-0.74</td>
<td>-0.71</td>
<td>-0.94</td>
<td>-0.88 **</td>
</tr>
<tr>
<td>No</td>
<td>0.86</td>
<td>2.14**</td>
<td>0.26</td>
<td>-1.59</td>
<td>-0.56</td>
<td>0.48**</td>
<td>0.64</td>
<td>0.47</td>
<td>-1.28**</td>
<td>-0.45</td>
<td>-0.58**</td>
<td>-1.31**</td>
<td>-1.67**</td>
<td>-2.00**</td>
<td>-1.75**</td>
</tr>
<tr>
<td>Missing</td>
<td>0.15</td>
<td>1.48</td>
<td>-0.53</td>
<td>-2.63**</td>
<td>-1.63</td>
<td>0.28</td>
<td>-0.39</td>
<td>0.07</td>
<td>-1.24</td>
<td>-1.50**</td>
<td>-1.90**</td>
<td>-3.73**</td>
<td>-3.53**</td>
<td>-2.73**</td>
<td>-3.45**</td>
</tr>
</tbody>
</table>

Population in the city where the high school is located (base category – less than 10,000)

| 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** | 3.06**| 2.27**  | 7.33**| 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** | 5.99**| 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   | -0.13 | -0.43 | -0.72 | -0.08  |

High school type (base – public school)

| Factor loadings | on $\theta_1$ (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 $\theta_2$ (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  | 0.35** | 2.24** | 1.00** |
|                | $\sigma^2$ | 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 5: Estimates of the GPA and the Score Equations.
The covariance between $\theta_q$ and $\theta_v$ 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. In contrast, the coefficients on prep school expenditure and school type are much higher. This makes sense as in Turkey, the best students are given scholarships to prep school as a way to attract paying customers. Moreover, Anatolian and Science schools are seen as the elite since they students admitted to them perform better both in terms of scores and GPA.

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 6 summarizes the explained variance that comes from observables, unobserved ability, and noise. These numbers are relatively stable across income groups with noise contributing only about 6% 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.

<table>
<thead>
<tr>
<th>Income</th>
<th>Observables, $X$</th>
<th>Unobserved ability, $\theta$</th>
<th>Noise, $\varepsilon_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>49%</td>
<td>45%</td>
<td>6%</td>
</tr>
<tr>
<td>Middle</td>
<td>47%</td>
<td>47%</td>
<td>6%</td>
</tr>
<tr>
<td>High</td>
<td>43%</td>
<td>51%</td>
<td>6%</td>
</tr>
</tbody>
</table>

Table 6: Contribution of Observables and Shocks

4.2 Results in Step 2

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

learn less while students in the middle income group learn the most. This is

\footnote{This suggests that the raw correlation between income and performance often seen in the data is being captured by our other controls.}
roughly in line with Frisancho 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 elite “exam” schools (where entry is competitive) versus public schools. Cumulative learning after three retakes is roughly of the same magnitude as the effect of being in an exam school.

Learning benefits the most middle income groups. Low income students get the second highest learning gains while the rich benefits the least from learning. This suggests that the less advantaged learn more when retaking. Moreover, we 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 school 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. To put these numbers into perspective, 25% of the first time retakers get more than 18 points on retaking, which is roughly the effect of going to an exam 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 banning retaking.

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 rich caring less about getting into the best schools. Since
their future success depends less on their exam performance due to better and more outside options (e.g. starting a business or joining the family firm) when compared to the poor, student who are better-off benefit little from moving from an already highly-ranked placement to a marginally better one.

5 Counterfactuals

We conduct a number of counterfactuals below, all aimed at reducing retaking. The 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 particular, we compare the no-retaking scenario (labeled as 1 attempt in Table 7) 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 7 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 look at the fraction of under placed students as it is the under placed who tend to retake. We define under placement 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 under placed students (where under placement is defined as being placed 5% below initial ability). Table 7 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 as the underplaced tend to retake. Learning shocks are an additional source of rank changes which makes retaking raise 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 of under placement limiting retaking raises under placement as expected.

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.
Policies: current – unlimited retaking, 1 attempt max, 2 attempts max, 5% penalty after attempt 1, the weight on GPA is doubled.

Table 7: Policy experiments

5.1 Limiting Retaking: who gains and loses?

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 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. 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

\footnote{By ex-post we mean that we keep the shocks faced by agents constant across policy scenarios.}
for. Taking these costs into account will reduce the number of losers ex-post.

Redistributional effects across income seem to arise mostly through differences in performance of different groups. 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.

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. We model utility as increasing in the score/rank of an agent. This would be true even if preferences were horizontal as a higher score makes more options available to a student.

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

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 under-

---

28 We 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.
standing of these tradeoffs is a topic for future work as the computation requirements would be considerable.

References


7 Appendix (For Online Publication)

7.1 Standardizing high school GPA.

The discussion below is based on that in Frisancho et. al (2013). Raw and stand-ardized 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, $A_{jk}$:

$$A_{jk} = \frac{\overline{GPA}_{jk}}{\overline{\text{Weighted Score}}_{jk}} \div \frac{\overline{GPA}_k}{\overline{\text{Weighted Score}}_k}$$ (7)

where $\overline{GPA}_{jk}$ and $\overline{\text{Weighted Score}}_{jk}$ are the average GPA and weighted scores for each high school and track combination. $\overline{GPA}_k$ and $\overline{\text{Weighted Score}}_k$ are the average GPA and weighted score across all comparable students from the same track.$^{29}$

The numerator in (7) 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 (7) 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:

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

$^{29}$This 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.
where $\overline{GPA}_{ijk}$ is defined as:

$$\overline{GPA}_{ijk} = \left( \frac{GPA_{ijk}}{A_{jk}} \right)$$

and $\overline{GPA}_{k}^{\text{max}}$ is just the maximum $\overline{GPA}_{ijk}$ in a given $k$. Notice that if the student is in a 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 (8) using the sample of first-time takers. We obtain the estimates of $\beta_g$ and $\beta_j$ by running each performance measure in (8) on $X$.

$$g = X'\beta_g + \theta'\alpha_g + \varepsilon_0$$

$$s_{j1} = X'\beta_j + \theta'\alpha_j + \varepsilon_{j1}$$

Then, we use the residuals from the above regressions to pin down the factor loadings $\alpha_g$ and $\alpha_j$, 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, $(\theta_v, \theta_q, \varepsilon_0, \varepsilon_{\text{math,1}}, \varepsilon_{\text{Turk,1}}, \varepsilon_{\text{ss,1}}, \varepsilon_{\text{sc,1}})$. 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 $\theta$ and the shocks $\varepsilon_0, \varepsilon_{\text{math,1}}, \varepsilon_{\text{Turk,1}}, \varepsilon_{\text{ss,1}}, \varepsilon_{\text{sc,1}}$ are independent of each other conditional on $X$.

Under the assumptions above one can identify $\alpha_g, \alpha_j$, the joint density of the common factors $[\theta_v, \theta_q]$, and the densities of the transitory shocks $\varepsilon_0$ and $\varepsilon_{j1}$ 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 $\theta$ 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 $\theta$.

Let $r$ be a vector of the five residuals from the system of equations (8)

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

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 $\theta_v$ and $\theta_q$ are expressed in terms of $\tilde{\theta}_v$ and $\tilde{\theta}_q$. Substituting for $\theta_v$ and $\theta_q$ 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 $\alpha_{\text{math}} = [0, 1]$ and $\alpha_{\text{Turk}} = [1, 0]$ respectively.

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

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 $I_\varepsilon$ is a diagonal matrix with the variances of $\varepsilon$’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 $\varepsilon$ shocks, two variances and one covariance of the common factors $\theta$, and six factor loadings $\alpha$. As we normalize the loadings in the math and Turkish equation to $\alpha_{\text{math}} = [0, 1]$ and $\alpha_{\text{Turk}} = [1, 0]$, four of the factor loadings in $\alpha$ 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 retaking threshold $c_1$ 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,

$$f_{s_2,g|a=2}(s_2, g) = \int \int f_{s_2,g,\varepsilon_1,\bar{s}_1|a=2}(s_2, g, \varepsilon_1, \bar{s}_1) d\varepsilon_1 d\bar{s}_1$$

$$= \int \int f_{\eta,\varepsilon_1,\bar{s}_1|a=2}(s_2 - \bar{s}_1, g, \varepsilon_1, \bar{s}_1) \left| \frac{\partial(\eta, g, \varepsilon_1, \bar{s}_1)}{\partial(s_2, g, \varepsilon_1, \bar{s}_1)} \right| d\varepsilon_1 d\bar{s}_1$$

$$= \int \int f_{g,\varepsilon_1,\bar{s}_1|a=2}(g, \varepsilon_1, \bar{s}_1) f_{\eta}(s_2 - \bar{s}_1) d\varepsilon_1 d\bar{s}_1$$

$$= \int \int f_{g,\varepsilon_1,\bar{s}_1|a=1}(g, \varepsilon_1, \bar{s}_1) I(\varepsilon_1 < c_1(\bar{s}_1)) f_{\eta}(s_2 - \bar{s}_1) d\varepsilon_1 d\bar{s}_1 \frac{1}{\Pr\{a = 2\}}$$

(9)
In the first line we integrate the joint distribution of second-time score, $s_2$, GPA, $g$, shocks $\varepsilon_1$ 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 $\eta$ to separate the marginal density of $\eta$ from the joint density of $\bar{s}_1, g$ and $\varepsilon_1$. This requires the following assumption:

**Assumption 2** The distribution of learning shocks, $\lambda_t$, and idiosyncratic shocks, $\varepsilon_t$, 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, $f_\eta$, one can predict the distribution of scores and GPA among the second-time takers. The estimates of $e_1$ and $f_\eta$ 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 9 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 $f_\eta$ (2 unknowns).

Note that one needs additional assumptions to separate learning, $\lambda_2$, from idiosyncratic shocks, $\varepsilon_2$, given the density of their sum, $f_\eta$. Assuming that $\varepsilon_2$ is drawn from the same distribution as $\varepsilon_1$ (identified from the factor model) and that $\varepsilon_2$ is orthogonal to $\lambda_2$, one can identify the distribution of $\lambda_2$ by deconvolution.

**Assumption 3** The distribution of idiosyncratic shocks $\varepsilon_t$ does not vary across attempts. Learning shocks $\lambda_t$ are independent of $\varepsilon_t$ 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 \( f_{s,g|a>4}(\bar{s}, g) \) 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,

\[
\begin{align*}
  f_{s,g|a>4}(\bar{s}, g) &= \left( f_{s,g|a=4}(\bar{s}, g) + f_{s,g|a>4}(\bar{s}, g) \right) F_{\xi}(e_4(\bar{s})).
\end{align*}
\]

Rearranging gives:

\[
\begin{align*}
  f_{s,g|a>4}(\bar{s}, g) &= \frac{f_{s,g|a=4}(\bar{s}, g) F_{\xi}(e_4(\bar{s}))}{1 - F_{\xi}(e_4(\bar{s}))}.
\end{align*}
\]

(10)

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

\[
\begin{align*}
  f_{s,g|a>4}(s, g) &= \int f_{s,g|a>4}(s - \varepsilon, g) f_{\xi}(\varepsilon) d\varepsilon \\
  &= \int f_{s,g|a>4}(\bar{s}, g) f_{\xi}(s - \bar{s}) d\bar{s} \\
  &= \int \frac{F_{\xi}(e_4(\bar{s}))}{1 - F_{\xi}(e_4(\bar{s}))} f_{s,g|a=4}(\bar{s}, g) f_{\xi}(s - \bar{s}) d\bar{s}
\end{align*}
\]

In the first line above we get the distribution of \( s = \bar{s} + \varepsilon \) from that of \( \bar{s} \) and \( \varepsilon \) using independence. In the third line we substitute using equation \( \text{(10)} \). The density \( f_{s,g|a=4} \) is identified as outlined above. Knowledge of \( F_{\xi} \) 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.