Gaming a Selective Admissions System

FRANCES XU LEE
Loyola University Chicago

WING SUEN
University of Hong Kong

January 29, 2022

Abstract. Costly manipulation to gain selective admission exhibits strategic complementarity when the admissions quota is loose but strategic substitution when the quota is tight. In a system with two layers of selection, gaming at the university entrance stage can induce a university to give preferential treatment to students from a selective high school, justifying why this school attracts better talent and causing gaming to unravel to the high school entrance stage. We apply this framework to evaluate the impacts of raising the university quota, abolishing university entrance examinations, eliminating sorting-by-ability in the high school system, and committing to low-powered selection policies.

Keywords: sequential manipulation; tutoring; ability sorting; equal credibility; preferential treatment

JEL Classification. D45; D82; J71

*We thank Jimmy Chan, Piotr Dworczak, Navin Kartik, Jin Li, Bruno Strulovici, and participants at the UNSW Organizational Economics Workshop and the Royal Economic Society conference for helpful comments and suggestions. This research is partly funded by the Research Grants Council of Hong Kong (Project No. 17500220).
1. Introduction

Standardized tests results are widely adopted as admissions criteria in education systems around the world. The reliance on test outcomes in the allocation of education resources is often motivated by the desire to select the most talented students to receive these resources. But high-stakes testing also provides powerful inducement for students and their families to engage in costly activities to boost test scores. Private tutoring for test preparation, for example, is prevalent in many countries, especially in Asia.\(^1\) In this paper, we use the terms tutoring, gaming, or manipulating interchangeably as a shorthand for all kinds of costly activities targeted at gaining an advantage in a selective admissions system. In the United States, students can fake disability status to allow them extra time or other accommodations when taking tests.\(^2\) Parents hire coaches to help with packaging their children’s college applications. The recent college admissions scandal (Korn and Levitz 2020) testifies to the extreme measures that some parents took to game the system. The goal of this paper is to develop a theoretical framework to study the incentives to engage in these gaming activities in a multi-stage setting, and their implications for selection outcomes.

Private tutoring is so common in Asia that education researchers dub it the “shadow education” system (Bray 1999; Lee, Park and Lee 2009). Although private tutoring may serve useful purposes such as helping the slower students catch up with the demands of formal schooling, the primary motivation behind tutoring seems to be to gain an advantage in competitive examinations. Bray and Kwok (2003) find that the proportion of secondary students in Hong Kong receiving tutoring increases as they reach higher grades. For students in Forms 6 and 7 (equivalent to senior high school), 85% of those receiving tutoring gave “examination preparation” as the main reason. In addition to students’ time input, the amount of financial resources that parents spend is huge. A Financial Times article estimates that Koreans spend $20 billion on private tutoring annually.\(^3\) Kaplan, a leading

---

\(^1\) Park et al. (2016) present data from Program for International Student Assessment surveys to show that the use of tutoring on any subject has increased substantially across a wide range of countries. Outside Asia, heavy users of tutoring (more than 40% of 15-year-olds in 2012) include students from Greece, Hungary, Poland, Russia, Spain, and Turkey. In the U.S., the use of tutoring is not particularly high compared to other countries, but the trend is increasing (from 12% in 2003 to 17% in 2012).

\(^2\) A federal disability designation known as 504 Plan allows students extra time for classroom assignments and standardized tests. Obtaining such a designation requires expensive psychological evaluation. According to a New York Times analysis (“Need Extra Time on Tests? It Helps to Have Cash,” July 30, 2019), the higher the income, the higher the share of high school students with this designation. In some rich communities, more than one in ten students have such a designation.

\(^3\) “South Korea’s Millionaire Tutors,” Financial Times, 16 June, 2014.
supplier of test preparation services in the U.S., has an annual revenue of $1.5 billion in 2018.

We treat gaming activities such as tutoring or buying disability designation as unproductive activities which work by improving students' observable test outcomes without affecting their true ability. Critics have long complained that drilling to prepare for standardized examinations and “teaching to the test” may be detrimental to education because they encourage rote learning, and important areas of knowledge are ignored if they are not on the test (e.g., Popham 2001; Volante 2004; see also Lazear 2006 for a different view). The *Financial Times* story mentioned earlier describes an English lesson in a Korean cram school, in which “little English is spoken in the lesson, which comprises an explanation of the TOEIC reading comprehension paper.” Some tutoring companies in China even systematically leaked test questions to their students.\(^4\)

Because these gaming activities work through distorting the information available to an admissions system, a main theme of our analysis is how informational externalities affect equilibrium level of gaming and equilibrium admissions outcome. We begin with a one-stage model in Section 2. Students can be either of high ability or low ability and they also differ on their costs of gaming the test. A university has limited capacity and prefers to select high-ability students. It does not directly observe students' abilities, but only their test scores, which can be improved by costly gaming. This section shows that tutoring can be strategic complements or strategic substitutes among students, but strategic substitution prevails when competition for university places is intense. This follows because the capacity limit causes diminishing returns to set in as many students get artificially high test scores through gaming. One implication is that a policy that aims to lessen the competition by increasing the university capacity can increase the payoff from gaming, and hence causes more students to engage in this activity.

We build on this model to consider in Section 3 an admissions system with two stages of selection—first at the high school level and then at the university level. Sorting-by-ability is a common feature in many education systems. In the U.S., for example, “magnet” high schools may select students based on entrance examinations. Under such a system, the university can rely on an additional signal, namely whether a student has got into a selective or non-selective high school, to guide its admissions decisions. The two-stage model brings out several interesting results.

---

First, though a selective high school is able to select a better pool of students, it also selects students with lower tutoring costs. As a result, at the university entrance stage, gaming incentives are higher (because manipulation is more credible when there are indeed more high-ability students) while the costs are lower, and students from the selective school will game the university entrance examination more intensely than those from the non-selective high school. Second, the greater extent of manipulation in the selective school neutralizes its initial advantage in the quality of its student pool, so that in equilibrium students who get admitted into university are equally likely to be high-ability students regardless of which high school they come from. In other words, high school affiliation is after all not informative to the university conditional on the second-stage test score, and the final selection outcome for the university only depends on the extent of gaming in the first stage. Third, despite the equilibrium uninformativeness of high school affiliation and the university’s sole objective of maximizing student quality, the university treats students from the selective school (who tend to be rich kids with low tutoring costs) more favorably. An outcome in which student quality is the same in both high schools and the university treats their students symmetrically is not stable, because small differences in the initial quality of the two schools would trigger competition to get into the school with the initial advantage, which in turn would amplify the initial difference in quality. The favorable treatment by the university creates a rent from entering the selective school and causes gaming to unravel to the first stage of the selection process.

Section 4 applies our framework to study a number of issues in selective admissions. Among the more interesting findings are:

1. Abolishing university entrance examinations would intensify gaming activity targeting the earlier tests (such as high school entrance examination or high school course grades), and as a result the average quality of the university’s admitted students would be lower.

2. When the cost structures of gaming the high school test and gaming the university test are the same, the university admits better students with two-stage selection than with one-stage selection. Therefore abolishing ability-sorting in the high school system worsens the university’s selection outcome. However, substantially more resources can be wasted on tutoring under ability-sorting.

3. The university may select better students by committing to a lower-powered admissions policy that is suboptimal ex post but that will mitigate the extent of gaming. Nevertheless the level of gaming under the optimal policy is positive in both stages,
and preferential treatment in favor of students from the selective school remains under the optimal policy.

Our analysis of tutoring and selective admissions is based on a model with very simple test technologies. In Section 5, we show that the main conclusions of our analysis are robust under various generalizations. Instead of assuming that tutoring is unproductive, we can allow it to be partially beneficial in transforming low-ability students into high-ability ones. Interestingly, a more beneficial tutoring technology may worsen the selection outcome for the university because it induces more tutoring, which obstructs information transmission. We provide a numeric example to illustrate this point in Section 5.1. In Section 5.2, we allow the high-ability students to also improve their test scores through tutoring and show that some high-ability students will pay for tutoring too as a result. Section 5.3 outlines a more general model with finitely many levels of student ability (instead of two levels), a continuum of test scores (instead of binary scores) and a test technology that is stochastic (instead of deterministic). This more general model can allow high schools to improve student abilities, with the selective school more able to do so than the non-selective school. We show that our basic results survive all of these extensions. In Section 6, we conclude with some observations concerning possible applications outside the context of college admissions.

Related Literature. Students' manipulation of test scores is a form of lying about their true ability, so our paper is related to the broad literature on costly lying (Kartik 2009; Kartik, Ottaviani and Squintani 2006). Chan, Li and Suen (2007) consider a signaling model of grade inflation, in which the cost of signaling is endogenously derived. Different from the existing literature, we focus on two-stage sequential gaming incentives, where gaming behavior depends on past outcomes, and the benefit from gaming unravels to an early stage. Moreover, the capacity constraint of the information receiver creates externality among the information senders.

Our paper shows that the university discriminates based on high school affiliation when there is ability-sorting in the high school system. This aspect is related to the large literature on statistical discrimination. Lundberg and Startz (1983) and Coate and Loury (1993) introduce models in which the labor market incorporates prior beliefs about the abilities of different groups of workers with their test scores to form estimates about their job qualification. In our model, group identity is not fixed, but is the result of first-stage of selection. We show that high schools with a larger pool of high-ability students and where students have lower tutoring costs are treated preferentially, despite the fact that
in equilibrium, conditional on the same university entrance test score, these students have the same expected ability as students from other schools.

There is a large literature on affirmative action, which is different from the notion of preferential treatment developed in this paper in a number of respects. First, preferential treatment in our model is the opposite of what “affirmative action” typically means: the elite school (where students are more advantaged) is being favored. Second, the affirmative action literature considers preferential treatment as a policy instrument, which assumes commitment ability to take ex post suboptimal actions. We study both the commitment and the no-commitment cases, and show that preferential treatment emerges in both cases. Third, a key component of our model is that tutoring effort is wasteful (at least partially), while most of the affirmative action literature assumes student efforts or investments are desirable. Fifth, the university in our model only wants to improve its selection outcome and does not care about diversity. One notable exception to the third and fourth points is Krishna and Tarasov (2016), who show that a university’s commitment to favor disadvantaged students can improve the selection outcome. Their model has two distinct student abilities: innate ability and acquired ability. Favoring students with low acquired ability tends to select students with high innate ability because the two types of abilities are substitutes in generating test performance. However, in our setup of no commitment, preferential treatment favoring the more advantaged students arises as an endogenous outcome. Our group identity is not only endogenous but is formed through the students’ strategic gaming in a sequential manipulation environment. In our setup, the ability to commit will reduce rather than enhance the degree of preferential treatment.

Our result on optimal admissions policy with commitment (Section 4.3) echoes the insights found in many different contexts that committing to underutilizing available data can mitigate harmful manipulation. Olszewski and Siegel (2019) also model wasteful test preparation in the context of college admission. They show that, with commitment, coarser test results can Pareto-improve the students’ payoffs through reducing wasteful gaming. Frankel and Kartik (2019) study signaling games in which agents have two-dimensional types: their academic ability and their ability to game the tests. They show that letting the students’ test scores be observed by more observers can worsen information for the existing observers. Ball (2021) and Frankel and Kartik (2021) adopt this framework to study design problems, and show how a decision maker can commit to underutilizing data.

---

to mitigate gaming and thereby improve allocation efficiency.

Because competition for limited spots in a university often takes the form of a contest, our research falls into the broad area of rent-seeking contests (Nitzan 1994; Tullock 2001). Siegel (2009) analyzes this type of interactions as an all-pay auction. Unique to our model is that the university does not directly care about how much investments students make in the contest; instead it is concerned about identifying true talents given that some low-ability students are investing to manipulate their test scores. Fang and Noe (2019) use an all-pay auction approach to study selection outcomes. Their focus is on risk-taking behavior in such contests. There is a small theoretical literature on cheating in rank-order tournaments (Curry and Mongrain 2009; Gilpatric 2010; Gilpatric and Reiser 2017). This literature studies how to deter cheating through the design of tournaments, audits, monitoring, and punishments in a single-stage setup. Our paper is focused on the information externalities in a two-stage contest that aims to learn about the private information of the participants.

2. One Stage of Selection

A large number of students compete for a limited number of university places. We model students as a continuum of agents and there are a unit mass of them. The total number (mass) of university places available is $Q \in (0, 1)$, which we refer to as the university quota.

A fraction $\lambda \in (0, 1)$ of the students have high ability. The remaining fraction have low ability. A student knows her true ability but others do not. Students have to sit an entrance examination to gain admission into university. All high-ability students get a high score $H$. Low-ability students get a low score $L$ if they do not pay for private tutoring. But if a low-ability student pays for costly tutoring, she gets score $H$ in the examination (without changing her true ability). Whether a student has paid for tutoring or not is not observable by the university. We present this simple model with binary types and binary deterministic examination scores in order to illustrate the logic of the analysis in the most transparent way. Section 5 extends the model in various directions, including allowing finitely many types and a continuum of stochastic scores.

A student gets benefit $B > 0$ from being admitted into the university. Since tutoring only affects a low-ability student’s score, only low-ability students will buy this service. Among these students, the tutoring cost $c$ is distributed according to distribution $F$. Assume $F$ has a density that is continuous and positive on $[0, B]$. A student’s payoff is the
benefit of getting into the university (if admitted), minus the cost of tutoring (if incurred). Because the net payoff is strictly decreasing in $c$, there is a cutoff cost level $S$ such that a low-ability student prefers to choose tutoring if and only if $c \leq S$. The mass of students who get tutoring is $(1 - \lambda)F(S)$.

Given the entrance examination scores, the university selects students to fill its available places. The university’s strategy is denoted $(X, Y)$, where $X$ is the probability of admission for high-score students, and $Y$ is the probability of admission for low-score students. We assume that the university prefers admitting a low-ability student to leaving its available spots unfilled. Since $\lambda + (1 - \lambda)F(S)$ is the total mass of high scorers, the university’s feasibility condition is:

$$(\lambda + (1 - \lambda)F(S))X + (1 - \lambda)(1 - F(S))Y = Q.$$  

The university’s objective is to maximize the average quality of its student intake, subject to filling its quota according to the feasibility constraint. We assume that the university cannot commit to a selection rule, so its admissions policy has to be optimal given the observed entrance examination results. We will re-visit the no commitment assumption in Section 4.3.

An equilibrium of this model requires: (1) Each low-ability student’s tutoring choice maximizes her payoff given the strategies of other students and admissions policy $(X, Y)$. (2) The university’s admissions policy is optimal given the observed examination scores and given the students’ strategies.

We refer to the posterior probability, given the test score, that a student is of high ability as her credibility. Because only low-ability students get a low score, the credibility of a low-scorer is 0. Let $K = \lambda/(\lambda + (1 - \lambda)F(S))$ denote the credibility of a high-scorer. It is optimal for the university to adopt a priority rule: high-scorers have a strictly higher priority of getting admitted (i.e., $X < 1$ implies $Y = 0$). If $Q > \lambda + (1 - \lambda)F(S)$, we say that the quota is loose, in which case all high-scorers get a place in the university $(X = 1)$ and the remaining quota is allocated randomly among low-scorers, which gives

$$Y = \frac{Q - (\lambda + (1 - \lambda)F(S))}{1 - (\lambda + (1 - \lambda)F(S))} > 0.$$  

---

6 This paper only considers a university whose objective is to select the highest-quality students. Selection on the basis of examination results is the dominant feature of many higher education systems in Asia. In other places a university may have other objectives such as diversity.
If $Q \leq \lambda + (1 - \lambda)F(S)$, we say that the quota is tight, in which case low-scorers do not get a place ($Y = 0$) and the quota is randomly allocated among high-scorers only, which gives

$$X = \frac{Q}{\lambda + (1 - \lambda)F(S)} \leq 1.$$  

Consider a low-ability student’s incentives. When the quota is loose, her probability of admission increases from $Y$ to 1 by acquiring tutoring. When the quota is tight, tutoring raises her probability of admission from 0 to $X$. Therefore her benefit from tutoring is:

$$\beta(S) = \begin{cases} 
B \left( 1 - \frac{Q - (\lambda + (1 - \lambda)F(S))}{1 - (\lambda + (1 - \lambda)F(S))} \right) & \text{if } Q > \lambda + (1 - \lambda)F(S), \\
B \left( \frac{Q}{\lambda + (1 - \lambda)F(S)} - 0 \right) & \text{if } Q \leq \lambda + (1 - \lambda)F(S).
\end{cases}$$

The benefit from tutoring increases in $S$ when the quota is loose and decreases in $S$ when the quota is tight. When $S$ is small, tutoring exhibits strategic complementarity: as more low-ability students get tutoring, it becomes increasingly difficult for low-scorers to gain admission into university, which makes tutoring more valuable. However, tutoring exhibits strategic substitution when the quota is tight: as a lot of students already obtain high scores in the examination through tutoring, the chance of admission for high-scorers falls, which lowers the benefit from tutoring. Many discussions on tutoring describe it as an “arms race,” in which parents send their children to tutoring schools because their neighbors are doing it.\footnote{Ramey and Ramey (2009) titled their paper on intensive parenting for college preparation, “The Rug Rat Race.” The Economist observes that, in Korea and other Asian economies, “Education has become an arms race in which one parent’s additional outlay of time and money forces others to follow suit.” (“Will Age Weaken the Asian Tiger Economies?” December 5, 2019.)} Our analysis suggests that the capacity constraint of a selective admissions system limits the extent of strategic complementarity. When the capacity of the university is reached, diminishing returns to tutoring set in as more and more students are getting high scores. Therefore the self-reinforcing tendency of this “arms race” eventually gives way to a self-limiting tendency to produce an equilibrium level of tutoring.

In the ensuing discussion we focus on the parameter case where the quota is tight. This reflects the view that competition for desirable university places is intense in many societies, so that obtaining high scores alone is not sufficient to guarantee admission. An assumption that will guarantee tight quota for any $S$ is $Q \leq \lambda$. In this case, $\beta(S)$ decreases in $S$. The following result is immediate, and the proof is omitted.
**Proposition 1.** Suppose $Q \leq \lambda$. The unique equilibrium tutoring cutoff $S^* \in (0, B)$ satisfies 
$$BQ/\left(\lambda + (1-\lambda)F(S^*)\right) = S^*.$$ 

The assumption $Q \leq \lambda$ used in Proposition 1 is sufficient but not necessary for equilibrium uniqueness. Consider the numerical example shown in Figure 1, where $\lambda = 0.25$, $B = 1$, and $F$ is a uniform distribution. In this example, equilibrium is unique when $Q < 0.81$ (left panel) and there are multiple equilibria when $Q > 0.81$ (right panel).

When multiple equilibria exist, the quota is tight at the largest equilibrium (i.e., the one with the highest value of $S^*$), because $\beta(S) = B > S$ at $Q = \lambda + (1-\lambda)F(S)$ and $\beta(S) < S$ at $S = 1$. In Figure 1, an increase in $Q$ shifts the $\beta(S)$ curve to the right. The value of $S^*$ in the largest equilibrium increases continuously as $Q$ increases. However, when $Q$ crosses the threshold value of (approximately) 0.81, two additional equilibria with smaller values of $S^*$ emerge, which correspond to the case of loose quota. The value of $S^*$ in the smallest equilibrium decreases as the quota expands beyond this threshold value. In this paper we focus on the largest (tight-quota) equilibrium because it always exists (for all values of quota) and it is locally stable in the sense of best-response dynamics.

**Corollary 1.** In the largest equilibrium, the amount of tutoring $S^*$ increases as the number of university places $Q$ increases.
Casual reasoning tends to suggest that the availability of more university places would soften the competition for admission and spur less tutoring. This reasoning is incomplete because it ignores the strategic substitution aspect of tutoring under a tight quota. When competition for admission is already quite intense, obtaining a high score in the entrance examination gives one a lottery for entering university but not a guarantee. The availability of more university places increases the chance of admission that comes with this lottery, and therefore raises the benefit of tutoring. The prediction of Corollary 1 is consistent with the experience of Hong Kong. The number of coveted public-funded universities has increased from two before 1991 to nine today. Interestingly, the tutoring industry boomed in Hong Kong during the 1990s and remains strong today.\(^8\) Gaining admission into university changed from a remote possibility into a realistic prospect, particularly if one is willing to spend resources to boost this prospect. Some commentaries in the U.S. tout increasing freshman class sizes at top universities as a way to combat gaming activities.\(^9\) Our analysis suggests that its effect can be the opposite.

3. Two Stages of Selection

In many education systems, selection occurs at multiple transition points of a student’s educational progression. For example, competition for admission into New York City’s nine specialized high schools is as intense, if not more, as competition for entry into universities, and it has spawned a cottage industry of tutoring to prepare for the Specialized High Schools Admissions Test. In some places this type of competition appears to have unraveled to earlier and earlier stages of education. A recent study (Chan et al. 2020) using boundary discontinuity design and cross-boundary matching strategy applied to housing prices suggests that Shanghai parents are willing to pay a 14% premium for an address that would secure the highest priority of getting their children into a local “superstar” primary school. In Hong Kong, a thriving business operates training classes to prepare toddlers for kindergarten interviews.\(^10\)

Building on the model in Section 2, we provide an analysis of the unraveling of tutoring for university admission to tutoring for high school admission in an education system with two stages of selection. The key to this analysis is that the university makes inference

---

\(^8\) Lui and Suen (2005) find that the college premium did not increase over this period, so there is little evidence for an increase in \(B\) that might have raised the demand for tutoring.

\(^9\) “Elite-College Admissions are Broken”, *The Atlantic*, October 14, 2018.

about student ability based on a student’s high school affiliation along with her university admission test score. High school affiliation provides the university information not only about the likely abilities of the students, but also about their tutoring cost because entry into different high schools are influenced by first-stage tutoring. Thus a student’s high school affiliation influences the university’s interpretation of her university admission test score based on how much the university test score is tainted by tutoring. There are strong interactions between the two stages of gaming.

To be sure, a selective high school can be different from non-selective ones in terms of the quality of its teachers, the amount of educational resources available, and a myriad of other factors that directly influence students’ educational experience and outcomes. We abstract from all these differences in order to emphasize the motive to gain an advantage in college admissions through the incentives to influence the university’s belief. Suppose School 1 is a selective high school and has a quota of \( q_1 \). School 2 is non-selective and has a quota of \( q_2 = 1 - q_1 \). All students who do not get into School 1 will attend School 2. For simplicity, we assume that neither high school changes its students’ ability. Let \( \lambda_1 \) and \( \lambda_2 \) be the proportion of high-ability students in School 1 and School 2, respectively. We have

\[
q_1 \lambda_1 + (1 - q_1) \lambda_2 = \lambda,
\]

where \( \lambda \) is the fraction of high-ability students in the population. In this model, which high school a student attends is a pure label that may or may not affect her chances of success at the next stage of selection by university. Section 5 briefly explains why the results will remain valid if the selective high school has an inherent advantage over the non-selective one in improving students’ ability.

Our model of competition for admission into School 1 is almost the same as that described in Section 2. Students know their own ability but School 1 does not. There is a high school entrance examination, the result of which is either a high score \( h \) or a low score \( l \) (we use lowercase letters to denote variables at the first stage of the selection process). High-ability students always get high score \( h \). Low-ability students get score \( l \) if they do not pay for tutoring, but will get score \( h \) if they pay.

The cost of tutoring (or gaming the admissions system in general) often depends on a student’s family background. Richer or more educated parents may incur lower costs due to a lower marginal utility of income or greater knowledge about the various options available in the educational system. We therefore expect that the cost of tutoring is persistent
between the two stages of selection. Let the “type” of a low-ability student be represented by \( c \). The cost of tutoring of a type-\( c \) student is \( c \) at the university entrance stage and is \( \delta c \) at the high school entrance stage. So \( \delta < 1 \) means that university selection is harder to game than high school selection, and \( \delta > 1 \) is the opposite case. The distribution of types in the population of low-ability students is given by \( F \) on \([0, B]\).

The timing of the two-stage game is as follows. In Stage 1, students decide whether or not to get tutoring. Then high school entrance examination scores realize. School 1 uses these scores to admit students. In Stage 2, students decide whether or not to get tutoring. Then college entrance examination scores realize. The university observes students’ high school affiliation and their college entrance examination scores, and uses this information to make admissions decisions. It does not observe students’ high school entrance examination scores.

Let \( s \) denote the cutoff type who is indifferent between getting tutoring or not in the first stage. School 1 aims to maximize the average ability of its students, subject to the constraint that it fills its entire quota \( q_1 \). We denote its admissions policy by \((x, y)\) where \( x \) is the probability of admission for students with score \( h \) and \( y \) is the probability of admission for students with score \( l \). Its feasibility constraint can be written as:

\[
(\lambda + (1-\lambda)F(s))x + (1-\lambda)(1-F(s))y = q_1.
\]

Students do not directly get any direct benefit from being in a particular high school. They get benefit \( B \) from being admitted into university.

For \( i = 1, 2 \), denote the cutoff type of low-ability student from School \( i \) who pays for tutoring in Stage 2 by \( C_i \). Let the university’s admissions policy be represented by \((X_1, X_2, Y_1, Y_2)\), where \( X_i \) is the probability that a high-scorer from School \( i \) is admitted, and \( Y_i \) is the probability that a low-scorer from School \( i \) is admitted. Let \( G_i(\cdot) \) denote the distribution of tutoring costs among low-ability students in School \( i \). The objective of the university is to maximize the average quality of the admitted students, subject to the feasibility condition:

\[
\sum_{i=1,2} q_i \left[ (\lambda_i + (1-\lambda_i)G_i(C_i))X_i + (1-\lambda_i)(1-G_i(C_i))Y_i \right] = Q.
\]

An equilibrium of this model requires: (1) Each low-ability student’s first and second-stage tutoring choices maximize her payoff given the tutoring choices of other students.
and given the admissions policies of School 1 and of the university. (2) The admissions policies of School 1 and of the university maximize the quality of their respective student intakes given the information available and given students’ tutoring choices.

**Assumption 1.** $Q < q_1 < \lambda$.

We assume that competition for admission into the selective high school and into university is intense. Assumption 1 is sufficient to guarantee that School 1’s quota is tight, implying that students with score $h$ in the high school entrance examination will be rationed (i.e., $x < 1$), while students with score $l$ will not have a chance of getting into School 1 ($y = 0$). Recall that $\lambda_i$ denote the proportion of high-ability students from School $i$ ($i = 1, 2$). By Bayes’ rule,

$$\lambda_1 = \frac{\lambda}{\lambda + (1 - \lambda)F(s)}, \quad (1)$$

and $\lambda_2 = (\lambda - q_1\lambda_1)/(1 - q_1)$. For any value of $s$, we have $\lambda_1 \geq \lambda_2$.

If the distribution of students in School 1 and School 2 were exogenous and unrelated to tutoring costs, we would have $G_1 = G_2 = F$. However, gaming induces disproportionately more students with low tutoring costs to successfully get into the selective School 1. Given that among low-ability students, only those with types lower than $s$ have a chance of attending School 1, we obtain

$$G_1(c) = \begin{cases} F(c) / F(s) & \text{for } c \in [0, s), \\ 1 & \text{for } c \in [s, B]. \end{cases}$$

The distribution of types in School 2 is determined by the identity:

$$q_1(1 - \lambda_1)G_1(c) + (1 - q_1)(1 - \lambda_2)G_2(c) = (1 - \lambda)F(c), \quad (2)$$

for any $c \in [0, B]$. Clearly, as long as some low-ability students choose not to get tutoring in Stage 1 (i.e., $s < B$), the distribution $G_2$ first-order stochastically dominates $G_1$.

Denote the credibility of a high-scorer (at the university entrance examination) from School $i$ ($i = 1, 2$) by $K_i$. The credibility of a low-scorer from either school is 0. By Assumption 1, it is not optimal for the university to allocate any of its available places to low-scorers; we have $Y_1 = Y_2 = 0$. We say high scorers from School 1 and School 2 are
equally credible if they are equally likely to be of a high ability, i.e.,

\[
\frac{\lambda_1}{\lambda_1 + (1 - \lambda_1)G_1(C_1)} = \frac{\lambda_2}{\lambda_2 + (1 - \lambda_2)G_2(C_2)}.
\] (3)

**Lemma 1.** In any second-stage subgame equilibrium with \(\lambda_1 > \lambda_2\), (a) high-scorers from School 1 and School 2 are equally credible; (b) the fraction of low-ability students who choose tutoring is strictly higher in School 1 than in School 2; and (c) high-scorers from School 1 are admitted with a weakly higher probability than those from School 2.

**Proof.** (a) Suppose \(K_i > K_j\), then high-scorers from School \(i\) have a higher priority of admission into university than high-scorers from School \(j\), i.e., \(X_i < 1\) implies \(X_j = 0\). If \(X_i < 1\), then \(X_j = 0\) would imply that no student in School \(j\) chooses tutoring. Therefore \(K_j = 1\), which contradicts \(K_i > K_j\). When \(X_i = 1\), all students in School \(i\) would choose tutoring. If \(i = 1\), \(X_1 = 1\) would imply that all students in School 1 are admitted into university, which violates \(Q < q_1\). If \(i = 2\), \(X_2 = 1\) would imply \(K_2 = \lambda_2 \leq \lambda_1 \leq K_1\), which contradicts \(K_2 > K_1\).

(b) Let \(\rho \equiv \lambda_1(1 - \lambda_2)/(\lambda_2(1 - \lambda_1))\). The equal credibility condition (3) can be written as \(\rho = G_1(C_1)/G_2(C_2)\). Therefore, \(\rho > 1\) implies \(G_1(C_1) > G_2(C_2)\).

(c) Since only students with types lower than \(s\) will choose tutoring and can enter School 1, the support of the type distribution \(G_1\) is \([0, s]\). For \(C_1 \leq s\), we have

\[
G_1(C_1) = \frac{F(C_1)}{F(s)} = \frac{\lambda_1(1 - \lambda)}{\lambda(1 - \lambda_1)}F(C_1),
\]

where the second equality uses equation (1). From the identity (2) and using the fact that \(q_1(1 - \lambda_1) = 1 - \lambda - (1 - q_1)(1 - \lambda_2)\), we can also obtain

\[
(1 - q_1)(1 - \lambda_2) \left(1 - \frac{G_2(C_1)}{G_1(C_1)}\right) = (1 - \lambda) \left(1 - \frac{F(C_1)}{G_1(C_1)}\right).
\]

This gives

\[
\frac{G_2(C_1)}{G_1(C_1)} = 1 - \frac{1 - \lambda}{(1 - q_1)(1 - \lambda_2)} \left(1 - \frac{\lambda_1(1 - \lambda_1)}{\lambda_1(1 - \lambda)}\right) = \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2)},
\] (4)

where the second equality uses the fact that \(\lambda = q_1\lambda_1 + (1 - q_1)\lambda_2\). Now, observe that the
equal credibility condition can also be written as:

\[
\frac{G_2(C_2)}{G_1(C_1)} - \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2)} = 0. \tag{5}
\]

Equations (4) and (5) together imply \( G_2(C_1) = G_2(C_2) \), and hence \( C_1 = C_2 \). If \( C_1 = C_2 < s \), the cutoff types must be indifferent between tutoring or not, and therefore \( BX_i = C_i \) for \( i = 1, 2 \) implies \( X_1 = X_2 \). If \( C_1 = C_2 = s \), low-ability students of all types in School 1 prefer tutoring to no tutoring. We have \( BX_1 \geq C_1 = C_2 = BX_2 \), and therefore \( X_1 \geq X_2 \).

The equal credibility result (Lemma 1(a)) is a consequence of an (ex post) optimal admissions policy, and it holds regardless of whether \( \lambda_1 \) and \( \lambda_2 \) are pre-determined or are the outcome of first-stage selection. One can think of the university as relying on two signals: high school affiliation and university test score. These two signals are sequential in the sense that students observe the outcome of first (high school affiliation) before deciding on whether to game the second (university test score). The equal credibility result says that, conditional on a positive realization of the second signal, the first signal is not informative to the university. Suppose this is not the case. Then, because of the limited capacity of the university, it will only select students with positive realizations of both signals. This would imply that students have no incentive to game the second signal if the first turns out to be negative, but will still have incentive to game the second signal if the first is positive. But then a negative realization of the first signal actually induces a better belief because the university is assured that the second signal is not gamed. This forms a contradiction.

Whenever there are more high-ability students in the selective school (i.e., \( \lambda_1 \geq \lambda_2 \)), gaming must also be more prevalent in the selective school in order for equal credibility to hold (Lemma 1(b)). If \( \lambda_1 \) and \( \lambda_2 \) are pre-determined and the distributions of types are identical in the two schools, \( G_1(C_1) \geq G_2(C_2) \) would immediately imply \( C_1 \geq C_2 \). When \( \lambda_1 \) and \( \lambda_2 \) are the outcome of first-stage selection, Lemma 1(c) shows that \( G_1(C_1) \geq G_2(C_2) \) still implies \( C_1 \geq C_2 \). Because the cutoff type of student who chooses tutoring is weakly higher in School 1 than in School 2, the indifference condition implies that the university must be admitting high-scorers in School 1 with a weakly higher probability than those in School 2.

Since high scorers from the two schools are equally credible, any admissions policy \( (X_1, X_2, 0, 0) \) is optimal as long as it is feasible. Nevertheless, Lemma 1(c) imposes the
restriction $X_1 \geq X_2$ on equilibrium admissions policies. This result holds in any subgame with an arbitrary first-stage cutoff type $s$. We can strengthen the weak inequality to a strict inequality by considering equilibrium incentives in the first-stage tutoring decision.

**Lemma 2.** In any equilibrium, $C_1 = C_2 = s \in (0, B)$ and $X_1 > X_2$.

**Proof.** If $s = 0$, then $\lambda_1 = 1$ and $K_1 = 1 > K_2$, which would violate the equal credibility condition. We therefore have $s > 0$. In the proof of Lemma 1(c), we show that $C_1 = C_2$. Suppose $C_1 = C_2 < s$. Then a type-$s$ student strictly prefers not to acquire second-stage tutoring no matter she is in School 1 or School 2. Because $Y_1 = Y_2 = 0$, the subgame payoff to type $s$ is 0. This implies that type $s$ strictly prefers not to acquire tutoring in the first stage, a contradiction. This shows that $C_1 = C_2 = s$. Lemma 1 establishes that $X_1 \geq X_2$. If $X_1 = X_2$, however, there is no gain from entering the selective high school, which would violate $s > 0$. We therefore must have $X_1 > X_2$. Finally, if $s = B$, all students in School 1 would have high score in the university entrance examination. This would imply $X_1 \leq Q/q_1 < 1$, and type $s$ would strictly prefer not to pay for tutoring. 

The highest type of low-ability student in School 1 is type $s$. Lemma 2 shows that all low-ability students in School 1 choose tutoring in the second stage, but only a fraction of low-ability students in School 2 do so. Despite sorting of high-ability students into the selective School 1, the second-stage gaming decisions neutralize the prior differences between the two schools ($\lambda_1 > \lambda_2$) to make high-scorers from both schools equally credible ($K_1 = K_2$). Nevertheless, any equilibrium must exhibit preferential treatment in favor of School 1, in that equally credible students are given a strictly higher chance of admission at the favored school ($X_1 > X_2$).

An equilibrium with $X_1 > X_2$ is supported by the fact that $1 = G_1(C_1) > G_2(C_2)$. In this equilibrium, type $s$ in School 1 strictly prefers to get tutoring, while type $s$ in School 2 is indifferent between tutoring or not. With $X_1 > X_2$, there is a rent of $B(X_1 - X_2)$ for both high-ability and low-ability students from being admitted into School 1. This rent provides an incentive for students to compete—by engaging in costly tutoring if necessary—for admission into the selective school, which in turn justifies why $\lambda_1 > \lambda_2$ can arise as an endogenous equilibrium outcome, despite the fact that one’s high school affiliation is just a pure label. The incentive to game the university admissions system unravels to the high school admissions stage through this channel.
Proposition 2. A unique equilibrium exists. Low-ability students with types lower than or equal to $s^*$ pay for tutoring at both stages, where

$$\left(\lambda + (1 - \lambda)F(s^*)\right)s^* = \frac{BQ}{1 + \delta}. \quad (6)$$

No one else pays for tutoring. The equilibrium admissions policy of the university is

$$(X_1^*, X_2^*, Y_1^*, Y_2^*) = \left(\frac{s^*}{B} + \frac{\delta Q}{(1 + \delta)q_1}, \frac{s^*}{B}, 0, 0\right). \quad (7)$$

Proof. Lemma 2 implies that type $s$ in School 2 is indifferent between tutoring or not. So we have $X_2 = s/B$. Since $Y_1 = Y_2 = 0$, the admissions probability $X_1$ is determined by the second-stage feasibility condition:

$$q_1(\lambda_1 + (1 - \lambda_1)G_1(C_1))X_1 + (1 - q_1)(\lambda_2 + (1 - \lambda_2)G_2(C_2))X_2 = Q.$$ 

Using the equal credibility condition (3) and the fact that $G_1(C_1) = 1$, this reduces to

$$X_1 = \frac{1}{q_1} \left( Q - (\lambda + (1 - \lambda)F(s) - q_1) \frac{s}{B} \right). \quad (8)$$

In the first stage, type $s$ can get into School 1 with probability $q_1/(\lambda + (1 - \lambda)F(s))$ by investing in first-stage tutoring at cost $\delta s$. The indifference condition requires:

$$\delta s = B(X_1 - X_2) \frac{q_1}{\lambda + (1 - \lambda)F(s)}.$$ 

Using the value of $X_1$ from equation (8) and $X_2 = s/B$, we obtain condition (6) stated in the proposition, and there is a unique $s^* \in (0, B)$ that satisfies this condition. Plugging $s = s^*$ into $X_1$ and $X_2$ gives the equilibrium admissions policy (7). 

The equilibrium condition (6) in Proposition 2 allows simple comparative statics analysis. As in Corollary 1, a larger university quota $Q$ leads to more gaming at both the university entrance stage and the high school entrance stage. A larger benefit from university admittance has the same effects. Similarly, suppose the cost of gaming falls in the sense of a first-order stochastic decrease in the distribution $F$. Equation (6) implies that the cutoff $s^*$ decreases, but the fraction of students choosing tutoring $F(s^*)$ increases. As
a result, the average quality of students admitted into university will fall.

Proposition 2 establishes that any equilibrium must exhibit preferential treatment in favor of the selective high school, despite the fact that high-scorers in both high schools are equally credible. This is because without preferential treatment, the more selective school will have a higher credibility. Preferential treatment induces more gaming in the selective high school to result in equal credibility between the schools.

The logic driving the preferential treatment outcome is different from the discrimination literature, which relies on multiple equilibria. In Coate and Loury (1993), for example, employers set different hiring standards for different groups, which produce different human capital investment incentives for these groups that in turn justify the different standards. In our model, low-ability students in the favored School 1 indeed invest more in tutoring (i.e., \( G_1(C_1) > G_2(C_2) \)), but such investments reduce rather than enhance credibility for the group as a whole. The preferential-treatment outcome is derived from costly gaming under the equal credibility requirement. Any equilibrium described in Proposition 2 necessarily exhibits discrimination in favor of students from School 1, whereas multiple equilibria in Coate and Loury (1993) is consistent with no discrimination if employers adopt the same standard for both groups. Furthermore, which high school a student belongs to is not exogenously given, but is the outcome of a selection process partly driven by competition for preferential treatment in the university entrance stage.

Our result that that there is a unique equilibrium exhibiting preferential treatment relies on the assumption that School 1 is selective while School 2 is not. Modifying this assumption can give rise to multiple equilibria, but the basic economic forces leading to preferential treatment does not change. Specifically, consider a model where both schools hold entrance examinations and try to recruit students with score \( h \). There are three equilibria: (1) The first equilibrium is identical to the one described in Proposition 2, with the university giving preferential treatment to students from School 1, and with tutoring unraveling to the first stage when students compete to get into this school. Despite the fact that School 2 also tries to select \( h \)-scorers, these students will choose to enroll in School 1 if offered admission there. (2) The second equilibrium is the flip side of the first, with School 2 being the high school which is sought after by students and which is accorded preferential treatment at the second stage. (3) The third equilibrium is one in which students randomly choose between the two schools if offered admission by both schools. The resulting distribution of student ability will be identical, and the equilibrium reduces to the one-stage model described in Proposition 1, with no incentive to get tutoring at the
high school entrance stage.

Note, however, that the third (symmetric) equilibrium is not robust in the following sense. Suppose the two high schools are not exactly identical; say, School 1 has a more beautiful campus that brings an additional payoff $\epsilon > 0$ to its students. No matter how small $\epsilon$ is, students who are offered admission by both schools will choose to enroll in School 1, destroying the symmetry in ability distribution between the two schools. On the other hand, the first equilibrium will survive if School 1 has an initial small advantage. To put it slightly differently, small differences in the ex ante quality of the two high schools can lead to intense competition for entry into the selective school that endogenously receives preferential treatment in equilibrium. More interestingly, the second equilibrium, in which School 2 is preferentially treated, will also survive even though it has a small initial disadvantage. This is because the informational externality arising from having a larger fraction of high-ability students is self-reinforcing: if $h$-scorers choose to join School 2 despite the uglier campus, preferential treatment by the university will produce a rent from joining School 2 that can overcome the payoff difference $\epsilon$ to make School 2 more attractive. Therefore, when both high schools can select students, endogenous selection will typically lead to asymmetric distribution of student ability. But which one ends up becoming the “elite school” may be affected more by history-dependent factors than by the underlying quality of the two schools.

4. Applications

4.1. Abolish standardized test requirements for university admission

A growing list of universities in the U.S., including Chicago, Harvard, and Princeton, have recently adopted test-optional admissions policies. The University of California system went further and decided in 2021 to stop considering SAT and ACT scores in its admissions decisions. The arguments for or against standardized tests are multi-faceted. We do not attempt to address all the relevant issues here, but will use the two-stage model of Section 3 to illustrate how abolishing university entrance examination may influence attempts at gaming admissions systems and their effects on selection outcomes. Because there is only one university in our model, considerations arising from competition among universities for high quality applicants are absent from this model.

We use a tilde to denote equilibrium quantities when there is no university entrance examination. Because of the lack of information from standardized test scores, the univer-
sity’s admissions policy is entirely determined by the high school affiliation of the students. When \( \tilde{\lambda}_1 > \tilde{\lambda}_2 \), the university gives students from School 1 strict priority over those from School 2, and so \( \tilde{X}_1 = Q/q_1 \) and \( \tilde{X}_2 = 0 \). This implies that the benefit of first-stage tutoring for a low-ability student is:

\[
\tilde{\beta}_1(\tilde{s}) = B \left( \frac{Q}{q_1} - 0 \right) \frac{q_1}{\lambda + (1 - \lambda)F(\tilde{s})}.
\]

The following result follows directly from the indifference condition \( \delta \tilde{s} = \tilde{\beta}_1(\tilde{s}) \), and its proof is omitted.

**Proposition 3.** Suppose the university entrance examination is abolished. In the unique equilibrium, all low-ability students with types less than or equal to \( \tilde{s} \) acquire tutoring, where

\[
(\lambda + (1 - \lambda)F(\tilde{s}))\tilde{s} = \frac{BQ}{\delta}.
\]

When there is no university examination, \( \tilde{X}_1 > X_1^* \) and \( \tilde{X}_2 < X_2^* \). Because the rent from entering School 1 increases from \( B(X_1^* - X_2^*) \) to \( B(\tilde{X}_1 - \tilde{X}_2) \), competition to get into the selective school becomes more intense, and we have \( \tilde{s} > s^* \). Without a university examination, equal credibility implies that the quality of the university’s intake is the same as the quality of an average student from School 1; we have \( \tilde{\lambda}_1 = \lambda/(\lambda + (1 - \lambda)F(\tilde{s})) \). With an entrance examination, equal credibility implies that the quality of the university’s intake is equal to \( K_1 \). Since \( G_1(C_1) = 1 \) in the equilibrium of Proposition 2, we have \( K_1 = \lambda_1 = \lambda/(\lambda + (1 - \lambda)F(s^*)) \). Therefore, \( \tilde{s} > s^* \) implies that the quality of the university’s intake is strictly lower without a university entrance examination.

**Corollary 2.** Abolishing the university entrance examination increases the amount of first-stage gaming and reduces the average ability of the university’s admitted students.

Interestingly, the drop in student quality is not directly due to the university losing a tool to measure student quality. The effect is indirectly caused by more gaming in an earlier stage to distort the remaining tool that the university relies more intensively on: a student’s high school affiliation.

### 4.2. No ability sorting in the high school system

Ability sorting via selective “magnet schools” is a common feature in the education systems of many localities. The merits and demerits of having selective high schools in the
system are often the subject of impassioned debates. Our simple model is not equipped to address the complex set of issues related to complementarity in human capital formation, peer effects, race, equity, or social mobility. Less remarked upon in these debates are the informational externalities conferred by elite schools, and the effects of these externalities on college admission outcomes. In this subsection, we use the analysis of Sections 2 and 3 to compare selection outcomes and the extent of gaming in education systems with and without ability sorting across high schools.

Take the one-stage selection model as representing a system where high schools are not differentiated in the ability mix of their student bodies. The average ability of a student admitted by the university is \( K = \frac{\lambda}{\lambda + (1-\lambda)F(S^*)} \), where \( S^* \) is the equilibrium level of tutoring described by Proposition 1. On the other hand, ability sorting can be represented by the two-stage model. By the equal credibility result, average ability of the university’s student intake with two stages of selection is not improved directly by sorting students into two high schools; it depends solely on the equilibrium level of gaming in the first stage: \( K_1 = \frac{\lambda}{\lambda + (1-\lambda)F(s^*)} \), where \( s^* \) is described by Proposition 2. Therefore the comparison of the quality of students admitted into the university boils down to a comparison between \( S^* \) and \( s^* \).

The conditions that pin down \( S^* \) and \( s^* \) are:

\[
S = \beta(S) = \frac{Q}{\lambda + (1-\lambda)F(S)}B, \\
\delta s = \beta_1(s) = \frac{q_1}{\lambda + (1-\lambda)F(s)}B\left[X_1^*(s) - X_2^*(s)\right].
\]

Proposition 2 shows that \( X_2^* > 0 \), which in turn implies \( X_1^* < Q/q_1 \) because the university allocates some of its quota to students from School 2. Therefore, we have \( \beta(S) > \beta_1(s) \) if \( S = s \). In a one-stage setup, gaming activities target the examination score that the university directly relies on. In a two-stage setup, the first stage of gaming targets at getting into the elite high school, but the university only partially relies on high school affiliation in making its admissions decision. This lower reliance reduces the benefit to game the admissions decision of the elite high school. If \( \delta = 1 \), the comparison of \( \beta(\cdot) \) against \( \beta_1(\cdot) \) immediately implies \( S^* > s^* \). However, this conclusion can be overturned if the cost of tutoring at the two stages are different, reflecting different levels of difficulties in gaming the two stages of selection. For example, the hourly rate for private tutoring at lower grades tends to be cheaper than that at higher grades. The following result summarizes
Proposition 4. There exists a critical value \( \hat{\delta} < 1 \) such that \( s^* > S^* \) and the university’s admitted students are of lower average ability under ability sorting (two stages of selection) than under no ability sorting (one stage of selection) if and only if \( \delta < \hat{\delta} \).

Proposition 4 states that ability sorting produces worse selection outcomes when the cost of first-stage tutoring is sufficiently lower than the cost of second-stage tutoring. Another relevant yardstick for comparison is the total expenditure spent on tutoring, which is a wasteful activity in this model. Under ability sorting, total expenditure (from both stages) on tutoring is:

\[
(1 + \delta) \int_0^{s^*} c \, dF(c).
\]

In contrast, total expenditure on tutoring with no ability sorting is

\[
\int_0^{S^*} c \, dF(c).
\]

If \( \delta \) is slightly above the critical value \( \hat{\delta} \), then \( s^* \) would be slightly below \( S^* \), so that ability sorting produces slightly better selection outcomes. But total resources spent on wasteful gaming would be substantially larger (almost by a factor \( 1 + \delta \)) under ability sorting.

4.3. Mitigate gaming with commitment

Thus far, we have assumed that the university cannot commit to an admissions policy. Its admissions decisions must be optimal ex post, given its available information and given the strategy of students. This means that the university will fully utilize the two available pieces of information: (1) students’ high school affiliation, and (2) university entrance examination scores. This gives students the incentive to manipulate both sources of information. In this section, we explore the possibility of using an ex post suboptimal admissions policy as a way of controlling the extent of manipulation. The general idea is that if high reliance on the two pieces of information encourages gaming targeted at them, then relying on them less (by adopting a low-powered selection scheme that differentiates students with different expected ability less) may reduce manipulation. To the extent that gaming reduces the efficacy of selection, the university may gain from low-powered selection even
if it does not care about the resources spent by students in gaming the selection system.\textsuperscript{11}

Viewed in this lens, the policy analyses in Sections 4.1 and 4.2 are special cases of commitments. Section 4.1 is about commitment to no reliance on second-stage test scores, while Section 4.2 is about commitment to no reliance on the high score affiliation. We have shown that both forms of commitment lead to worse selection outcomes (if $\delta > \hat{\delta}$). However, an optimal scheme in which the university can fine-tune its reliance on each piece of information in its admissions may produce strictly better outcomes.

Consider the one-stage model where the university commits to admissions policy $(X, Y)$. The university’s problem is:

$$\max_{X,Y,S} \quad X$$

subject to

$$(\lambda + (1 - \lambda)F(S))X + (1 - \lambda)(1 - F(S))Y = Q,$$

$$B(X - Y) = S,$$

$$Y \geq 0.$$

High-ability students have score $H$, and the objective is to maximize their chance of admission $X$. The constraint $B(X - Y) = S$ reflects the response of low-ability students to game the admissions system through costly tutoring. Note also that $Y \geq 0$ holds if and only if $S \leq S^*$ (where $S^*$ is the equilibrium cutoff in the one-stage model of Section 2).

Substituting out the variables $X$ and $Y$ reduces this problem to:

$$\max_{S \leq S^*} \quad Q + (1 - \lambda)(1 - F(S))\frac{S}{B}.$$

If we introduce the assumption that the distribution $F$ satisfies the (weak) monotone hazard rate property, then $(1 - F(S))S$ is quasi-concave in $S$, and is strictly concave whenever it is increasing. Denote

$$\hat{S} \equiv \arg\max_S \ (1 - F(S))S.$$

If $\hat{S} \geq S^*$, then the solution to the commitment problem is $S = S^*$, and the optimal commitment policy coincides with the equilibrium no-commitment policy. If $\hat{S} < S^*$, 

\textsuperscript{11}Frankel and Kartik (2021) and Ball (2021) show that a decision maker should commit to underutilize data to reduce manipulation. Whitmeyer (2021) shows that an information receiver can benefit from committing ex ante to observe a noisy signal of the message from the sender. In the finance literature, Goldman and Slezak (2006) point out that high-powered stock option incentives may induce manipulation.
then the optimal solution is $S = \hat{S}$. The corresponding admissions policy is given by $Y = Q - (\lambda + (1 - \lambda)F(\hat{S}))\hat{S}/B$ and $X = Y + \hat{S}/B$. Note that $\hat{S} < S^*$ implies $X - Y < X^* - Y^*$; thus the optimal policy is lower-powered compared to the equilibrium policy. Moreover, since $Y^* = 0$, $X - Y < X^* - Y^*$ implies that $Y > 0$. That is, the optimal policy does not follow a strict priority rule; students with score $L$ have positive probability of getting into university even though they are known to have low ability. Finally, note that choosing $S = 0$ is not optimal even if the university has commitment power. To eliminate all gaming incentives, the university would have to place no reliance on test scores, which would waste valuable information on student ability.

Recall that $S^*$ increases with the size of the quota $Q$ (Corollary 1), while $\hat{S}$ does not vary with $Q$. Given the assumption that $F$ satisfies the monotone hazard rate property, there exists a critical quota $\hat{Q}$ such that $\hat{S} < S^*$ if and only if $Q > \hat{Q}$. This suggests that commitment ability strictly improves the university’s selection outcome if its quota is sufficiently large.

The same logic extends to the two-stage model. Let the university commit to a policy $(X_1, X_2, Y_1, Y_2)$. Proposition 5 below shows that the university’s optimal commitment policy induces students with types below $\hat{s}$ to acquire tutoring at both stages, regardless of whether they are admitted to School 1 or School 2, where

$$\hat{s} = \begin{cases} s^* & \text{if } \hat{S} \geq s^*, \\ \hat{S} & \text{if } \hat{S} < s^*, \end{cases}$$

and $s^*$ is the equilibrium cutoff for the first-stage tutoring decision in Section 3. Whenever $\hat{S} < s^*$ (which occurs when $Q$ is large), the optimal commitment scheme does not follow a priority rule.

**Proposition 5.** Suppose $F$ satisfies the monotone hazard property. The university’s optimal commitment policy induces the tutoring cutoffs, $C_1 = C_2 = s = \hat{s}$, and is given by

$$Y_2 = Q - \frac{1 + \delta}{B}(\lambda + (1 - \lambda)F(\hat{s}))\hat{s},$$

and,

$$X_1 = Y_2 + \frac{\hat{s}}{B} + \frac{\delta(Y_2 - Y_1)}{(1 + \delta)q_1}, \quad X_2 = Y_2 + \frac{\hat{s}}{B}, \quad Y_1 \in \left[0, X_1 - \frac{\hat{s}}{B}\right].$$

The optimal policy does not follow the strict priority rule (with $Y_2 > 0$) if and only if $\hat{s} < s^*$,
and it coincides with the no-commitment equilibrium policy (with $Y_2 = 0$) if and only if $\hat{s} = s^*$.

In the proof of Proposition 5 (in Appendix A) we first show that a policy that induces no tutoring ($s = 0$) is dominated by a policy that induces the no-commitment equilibrium allocation. Hence the optimal policy entails $s > 0$. We then use the monotone hazard property which ensures that $(1 - F(s))s$ is strictly concave when it is increasing, together with Jensen’s inequality, to establish that whenever there is a feasible policy such that $C_1$, $C_2$ and $s$ are not all equal, there exists another feasible policy with $C_1 = C_2 = s$ which will strictly improve the quality of the university intake. The problem then reduces to finding the value of $s$ that maximizes $(1 - F(s))s$.

Under the optimal commitment policy, a low-ability student in School 2 is indifferent between tutoring and no tutoring if her type is $\hat{s} = B(X_2 - Y_2)$. So $\hat{s} < s^*$ will imply that the premium attached to a high test score $H$ (i.e., $X_2 - Y_2$) is lower under optimal commitment than under no commitment. Similarly a low-ability student is indifferent between first-stage tutoring and no tutoring if her type satisfies $\delta \hat{s}(\lambda + (1 - \lambda)F(\hat{s})) = B(X_1 - X_2)q_1$. Again, $\hat{s} < s^*$ will imply that the preferential treatment attached to affiliation with the selective high school (i.e., $X_1 - X_2$) is lower under optimal commitment than under no commitment.

A lower-powered selection system improves selection outcomes by mitigating the extent of gaming activities. Lowering $X_2 - Y_2$ and $X_1 - X_2$ entails raising $Y_2$ for the quota constraint to be satisfied. This explains why $Y_2 > 0$ under the optimal commitment policy whenever $\hat{s} < s^*$.

Adhering to an ex post suboptimal admissions rule may be difficult. A practical way for the university to achieve the commitment outcome is through the adoption of a low-powered test technology. Consider the one-stage model and let the original test technology be denoted by $\mathcal{E}$, with scores $H$ or $L$ as already described. Consider a less informative test technology $\mathcal{E}'$ which is a garbling of $\mathcal{E}$. A student with score $H$ under $\mathcal{E}$ gets score $H'$ under $\mathcal{E}'$ with probability 1. A student with score $L$ under $\mathcal{E}$ gets score $H'$ or $L'$ (with probabilities $z$ and $1 - z$) under $\mathcal{E}'$. Given some assumption about the costs of manipulation under different test technologies, the outcome of the optimal commitment policy $(X, Y)$ under

---

12 The reason why $Y_1$ can be positive is different. Under the optimal commitment policy, all low-ability students in School 1 prefer second-stage tutoring to no tutoring as long as $Y_1 \leq X_1 - \hat{s}/B$. For any $Y_1$ that satisfies this restriction, no student in School 1 will get score $L$, and therefore the precise value of $Y_1$ does not affect the selection outcome.

13 In Chan and Eyster (2003), the use of non-cutoff admissions policy stems from a preference for diversity and is part of the university’s optimal response to a ban on affirmative action.
can be implemented by an ex post optimal policy \((X', Y')\) under \(\mathcal{E}'\) with an appropriate choice of \(z\). Specifically, suppose the cost of changing score \(L'\) to \(H'\) under test technology \(\mathcal{E}'\) for a low-ability student of cost type \(c\) is \(\kappa c\). Let \(z = X/Y\), and let \(X' = \Pr[\text{admit } | H'] = X\) and \(Y' = \Pr[\text{admit } | L'] = 0\). Under test technology \(\mathcal{E}'\) and admissions policy \((X', Y')\), the net benefit from tutoring for a type-\(c\) student is \(B(1-Y/X)(X-0)-\kappa c\). If \(\kappa = 1\), this net benefit is the same as that under test technology \(\mathcal{E}\) and admissions policy \((X, Y)\). Moreover the quota constraint is satisfied as long as the original policy \((X, Y)\) is feasible. Thus the less informative test technology \(\mathcal{E}'\) with ex post optimal admissions policy \((X', Y')\) can replicate the optimal commitment outcome under \(\mathcal{E}\), provided that the cost of manipulation is the same. However, if it is cheaper to manipulate a less informative test, and in particular if \(\kappa = 1-z\), adopting the less informative test technology \(\mathcal{E}'\) together with an ex post optimal admissions policy will simply replicate the original equilibrium outcome under \(\mathcal{E}\) with no commitment.

In the two-stage model, even if we assume \(\kappa = 1\), it is not true that the optimal commitment outcome under \(\mathcal{E}\) can be implemented by an ex post optimal admissions policy under a less informative test \(\mathcal{E}'\) for an appropriate choice of \(z\). To illustrate this point, suppose we choose \(z = Y_2/X_2\) and pick \((X'_1, X'_2, Y'_1, Y'_2) = (X_1, X_2, 0, 0)\), where \(X_1, X_2\), and \(Y_2\) are given by Proposition 5. It can be verified that the optimal commitment outcome can be replicated. However, the policy \((X'_1, X'_2, Y'_1, Y'_2)\) is not ex post optimal. Because students with score \(H'\) in School 1 are more credible than those with score \(H'\) in School 2, the university could gain ex post by admitting the former set of students with a strict priority over the latter. Although adopting a less informative test \(\mathcal{E}'\) cannot replicate the commitment outcome, it will generally improve the quality of the university’s intake if we maintain the assumption that \(\kappa = 1\) for \(z > 0\). This assumption amounts to keeping the test technology \(\mathcal{E}\) unchanged and raising the cost of gaming by a factor \(1/(1-z)\). Not surprisingly, increasing \(z\) will improve selection outcomes because low-ability students would have less incentive to manipulate test scores if they already can get score \(H'\) with positive probability without gaming.

\footnote{The proof of Lemma 1(c) shows that \(C_1 = C_2\) implies the equal credibility condition. Since the optimal commitment policy for test \(\mathcal{E}\) induces \(C_1 = C_2\), \(H\)-scorers from both schools are equally credible. Under policy \((X'_1, X'_2, Y'_1, Y'_2)\) for the alternative test technology \(\mathcal{E}'\), the mass of students who get score \(H'\) in School 1 is the same but the mass of students who get score \(H'\) in School 2 will be higher. On the other hand, the mass of high-ability students in the two schools remains unchanged. Therefore, those with score \(H'\) in School 2 are less likely to be high-ability students compared to those with score \(H'\) in School 1.}
5. Discussion

5.1. Tutoring may improve student ability

In this section, we relax the assumption that tutoring is unproductive. Specifically, suppose in addition to changing the high school or university test scores, each stage of tutoring also transforms a low-ability student to a high-ability one with probability \( \omega \in [0, 1) \). We keep assuming that a student’s payoff is the benefit of getting into the university minus the cost of tutoring, if incurred. In other words, students do not directly care about having high or low ability.

In the two-stage setup, the equal credibility condition still holds. First, the university has to take in a high-scorer from any school with a positive probability (i.e., \( X_1, X_2 > 0 \)), because otherwise that school will have no tutoring and its credibility will be perfect, which is contradictory to not admitting any high-scorer from that school. Second, \( X_1 < 1 \), because it will otherwise violate the university’s quota (Assumption 1). Third, \( X_2 < 1 \). If \( X_2 = 1 \), there is no incentive to get tutoring to get into School 1, which implies perfect credibility for School 1’s high-scorers, and this will contradict favoring School 2 in college admission. This argument establishes that \( X_i \in (0, 1) \) for \( i = 1, 2 \), which can only happen when \( K_1 = K_2 \):

\[
\frac{\lambda_1 + (1 - \lambda_1)G_1(C_1)\omega}{\lambda_1 + (1 - \lambda_1)G_1(C_1)} = \frac{\lambda_2 + (1 - \lambda_2)G_2(C_2)\omega}{\lambda_2 + (1 - \lambda_2)G_2(C_2)}.
\]

This simplifies to the same equal credibility condition (5) as in the base setup with \( \omega = 0 \). Therefore, \( \lambda_1 > \lambda_2 \) implies \( G_1(C_1) > G_2(C_2) \). That is, gaming college admissions is more prevalent at the elite high school than at the non-elite school.

Thus, parts (a) and (b) of Lemma 1 are still valid when \( \omega > 0 \). In Appendix B, we also show that equal credibility implies \( C_1 = C_2 \), as required by Lemma 1(c). Unlike the case for \( \omega = 0 \), however, we can either have \( C_1 = C_2 = s \) or \( C_1 = C_2 < s \).

**Proposition 6.** For any \( \omega \in [0, 1) \), there is a unique cutoff value \( s^*(\omega) \) such that:

(a) if \( \omega \leq \delta \), the unique equilibrium is characterized by the cutoffs \( C_1 = C_2 = s = s^*(\omega) \), with \( s^*(0) = s^* \) (given in Proposition 2);

(b) if \( \omega > \delta \), the unique equilibrium is characterized by the cutoffs \( s = s^*(\omega) \) and \( C_1 = C_2 = (\delta/\omega)s^*(\omega) \), with \( s^*(\omega) > S^* \) (given in Proposition 1).

Furthermore, \( s^*(\omega) \) increases in \( \omega \).
The proof is provided in Appendix B. Note that this proposition encompasses Proposition 2 as a special case when \(\omega = 0\). There are two interesting observations derived from Proposition 6. First, unlike the case where tutoring is unproductive, it is possible to have \(C_1 = C_2 < s\) if \(\omega > \delta\). A type-\(s\) student chooses tutoring at the first stage but chooses no tutoring at the second stage. It is worth choosing first-stage tutoring if the cost is low enough (\(\delta\) is small), and if the probability that she will turn into a high-ability student is high enough (\(\omega\) is large) so that she will have a chance of getting into university even without second-stage tutoring.

The second observation is that \(s^*(\omega)\) increases in \(\omega\). To be sure, this is not a surprising result, as more productive tutoring increases its benefit. But this observation has an interesting implication for the university’s selection outcome. For illustration, suppose that \(\omega \leq \delta\), so that \(C_1 = C_2 = s = s^*(\omega)\). Because all students in School 1 choose tutoring, the equal credibility condition implies that the equilibrium credibility of an admitted student in the university is:

\[
K^* = \lambda_1 + (1 - \lambda_1)\omega = \frac{\lambda + (1 - \lambda)F(s^*(\omega))\omega(2 - \omega)}{\lambda + (1 - \lambda)F(s^*(\omega))}. \tag{9}
\]

An increase in \(\omega\) has a direct positive effect on \(K^*\), as tutoring turns more students into truly high-ability students. But an increase in \(\omega\) also raises the fraction of students who choose to manipulate test scores, and this will hurt the equilibrium credibility of admitted students. The overall effect is therefore ambiguous. For example, for \(\omega \leq \delta\), Appendix B shows that \(s^*(\omega)\) is given by the solution to:

\[
(\lambda + (1 - \lambda)F(s))s = \frac{BQ}{1 + \delta - \omega}.
\]

Let \(B = 1\), \(\delta = 0.1\), \(\lambda = 0.2\), \(Q = 0.1\), and \(F(x) = x^4\). When tutoring is unproductive (i.e., \(\omega = 0\)), we have \(s^*(0) = 0.4088\). The equilibrium credibility is \(K^* = 0.8995\). For \(\omega = 0.05\), we have \(s^*(0.05) = 0.4224\), and the corresponding credibility is \(K^* = 0.8981\). In this example, an increase in productivity of tutoring reduces the average quality of an admitted student, even though it increases the mass of high-ability students in the application pool.

For fixed \(\omega\), equation (9) shows that the credibility of the students admitted into the university, \(K^*\), is strictly decreasing in the extent of first-stage gaming \(s^*\). Therefore, the insights of Sections 4.1 and 4.2 on the applications remain valid.
5.2. High-ability students may also pay for tutoring

In our base model, we assume that only low-ability students can improve test scores through tutoring. In this section, we show that the results extend when high-ability students can also improve test scores through tutoring.

We modify the setup by assuming that a high-ability student has probability \( \pi \in (0, 1] \) of getting \( H \) score without tutoring, but will get \( H \) score for sure with tutoring. We assume that a low-ability student’s tutoring allows her to mimic a high-ability student, which means tutoring increases her probability of getting \( H \) score from 0 to \( \pi \). The base models in Section 2 and 3 correspond to the case of \( \pi = 1 \). Notice that if \( \pi > 0.5 \), then a low-ability student gains more from tutoring than a high-ability student.

Let \( S' \) denote the cutoff tutoring cost for a high-ability student in the one-stage setup. We continue to use \( S \) to denote the cutoff tutoring cost of a low-ability student. The university’s feasibility condition is:

\[
\lambda(F(S') + (1 - F(S'))\pi) + (1 - \lambda)F(S)\pi \]X
\[+ \left[ \lambda(1 - F(S'))(1 - \pi) + (1 - \lambda)(F(S)(1 - \pi) + (1 - \lambda)(1 - F(S))) \right] Y = Q.
\]

A sufficient condition for the university quota to be tight is \( Q < \lambda \pi \), which we maintain for this section. Denote the credibility of a high-scorer by \( K^H \) and that of a low-scorer by \( K^L \) (which may not be zero). Observe that it is impossible to have an equilibrium with \( F(S) = 1 \), because tight quota implies that even a high-scorer cannot be admitted to university with probability 1. After ruling out this case, it is clear that high-ability students have a strictly higher chance of getting \( H \) score than low-ability students regardless of tutoring choices: \( F(S') + (1 - F(S'))\pi = \pi + F(S')(1 - \pi) > F(S)\pi \) for any \( \pi \in (0, 1] \). Therefore, \( K^H > K^L \), which implies that the university adopts a priority rule and allocates its quota randomly among high-scorers.

Once we determine that \( Y = 0 \), we can use the feasibility constraint to pin down \( X \) as a function of \( S \) and \( S' \). The cutoff type of low-ability students satisfies \( S^* = \pi BX \). The equilibrium value of \( S^* \) satisfies

\[
\frac{\pi BQ}{\lambda(F(S^*) + (1 - F(S^*))\pi) + (1 - \lambda)F(S^*)\pi} = S^*,
\]

with \( S'^* = (1 - \pi)S^*/\pi \) because the benefit from tutoring for a high-ability student is
\[(1 - \pi)/\pi\] times that for a low-ability student. The above equation also shows that the benefits from tutoring for the low-ability and high-ability students are decreasing in \(S^*\) and \(S'^*\). Hence an increase in university quota \(Q\) raises the equilibrium extent of tutoring \(S^*\) and \(S'^*\), just as predicted by the base model.

A similar analysis applies to the case of a two-stage selection model with \(\pi \in (0, 1]\). The details of that analysis is relegated to the Online Appendix. We show that equal credibility continues to hold. Those who have paid for tutoring in the first stage go on to pay for tutoring in the second stage regardless of which high school they attend. Moreover, high-scorers from School 1 are treated preferentially over high-scorers from School 2 by the university. The main results in our base model are not affected by allowing high-ability students to also choose to manipulate their test scores.

### 5.3. Multiple types and continuous scores

We develop two key insights in this paper concerning gaming in college admissions. The first point is that gaming can exhibit both strategic complementarity and strategic substitution, with strategic substitution prevailing when the admissions quota is tight. This leads to the prediction that increasing the number of college places can increase the equilibrium level of tutoring. The second point relates to the unraveling of gaming incentives to earlier stages of the educational progression, leading to competition for entry into selective high schools or even if they are not superior in intrinsic quality. These points are illustrated in the context of a very simple model with binary ability distribution, binary test scores, and a deterministic gaming technology that always produces a high score for low-ability students. In this section, we give a brief discussion of a more general model environment that delivers similar insights. Details of this model are provided in the Online Appendix.

Consider the following model environment:

- The ability distribution allows multiple types of student abilities, \(a_1 < \ldots < a_N\).
- The test technology is stochastic and test scores are continuously distributed. A student with ability \(a_n\) obtains test score \(T = a_n + U\), where \(U\) is a continuous random variable distributed according to \(\Phi\) with a log-concave density \(\phi\).
- A student with ability \(a_n\) (for \(n < N\)) gets test score \(T = a_{n+1} + U\) by investing in tutoring.
- The tutoring cost distribution can depend on student ability type in an arbitrary way.

Equilibrium in this more general environment is characterized by a cutoff test score \(\hat{T}\).
to satisfy the university’s quota constraint. The incentive for a student with ability $a_n$ (for $n < N$) to get tutoring is

$$\beta(a_n, \hat{T}) = B \left( \Phi(\hat{T} - a_n) - \Phi(\hat{T} - a_{n+1}) \right).$$

The relevant observation is that $\beta(a_n, \hat{T})$ is increasing then decreasing in $\hat{T}$ under the log-concavity assumption. When the quota is loose and the admissions threshold $\hat{T}$ is low, a greater extent of gaming would raise the admissions standard and increase the benefit from gaming. However, the opposite effect obtains when the quota is tight and $\hat{T}$ is already high. In this case a greater extent of gaming would further raise the admissions standard and reduce the benefit from gaming (i.e., strategic substitution). This reflects the fact that it becomes increasingly difficult to gain an advantage in meeting the threshold as the standard becomes very high.

Working with this general model in the two-stage setup is challenging, because keeping track of the progression of $N$ different types of students is analytically difficult. Nevertheless we can still obtain the unraveling result if we let $N = 2$ while maintaining the continuous-scores test technology. The university’s admissions policy can be summarized by two test score thresholds, $(T_1, T_2)$, for students from Schools 1 and 2. The optimal admissions policy requires the university to equalize the credibility of the marginal student from the two schools, which can be expressed (in odds ratio form) as:

$$\frac{\lambda_1}{1 - \lambda_1 G_1(C_1)\ell(T_1) + 1 - G_1(C_1)} = \frac{\lambda_2}{1 - \lambda_2 G_2(C_2)\ell(T_2) + 1 - G_2(C_2)},$$

where $\ell(T_i) = \phi(T_i-a_2)/\phi(T_i-a_1)$. In the Online Appendix, we show that $\lambda_1 > \lambda_2$ implies $T_1 < T_2$. Intuitively, if there were no gaming of test scores, standard Bayesian reasoning suggests that the university should pay attention to both prior knowledge about the quality of the high schools and information from test scores, and therefore $\lambda_1 > \lambda_2$ would imply $T_1 < T_2$. The logic of this result has been pointed out, for example, in the literature on optimal judicial standards (Farmer and Terrell 2001). We show that this result survives even when students endogenously choose the extent of manipulation.

The fact that $T_1 < T_2$ means students from School 1 are treated preferentially by the university. The rent from getting into School 1 creates an incentive for students to spend costly resources to get into this school, causing tutoring to unravel to the first stage, while in turn justifying why School 1 will have better students than the other school. The logic
is the same as described in our main model.

We also note that the key result, \( \lambda_1 > \lambda_2 \) implies \( T_1 < T_2 \), still holds if we let School 1 to be better at bringing some inherent value to the students. For example, suppose School 1 transforms a share \( \zeta_1 \) of its low-ability students into high-ability ones, while School 2 transforms a share \( \zeta_2 \) of its low-ability students into high-ability ones, with \( 1 > \zeta_1 \geq \zeta_2 \geq 0 \). The transformation is independent of the cost type of the students, and the students who gain a higher ability are aware of it before deciding whether to game the university test. With this modification, equal credibility still holds, but now the credibility of high-scorers is calculated based on the post-transformation proportion of high-ability students in each school, instead of the initial proportion. Let \( \hat{\lambda}_i \) \((i = 1, 2)\) denote the proportion of high-ability students in School \( i \) after the transformation has happened. Given the assumptions, we immediately have \( \lambda_1 > \lambda_2 \) implies \( \hat{\lambda}_1 > \hat{\lambda}_2 \), and the proof of the key result in the Online Appendix goes through with \( \lambda_i \) replaced by \( \hat{\lambda}_i \).

6. Conclusion

Information manipulation is common in many aspects of life when the information collected has implications for resource allocation. Creative accounting, fake product reviews, and credit score management are among the more important examples. In this paper we focus on information manipulation in college admissions, partly because of its significance to students and their families, and partly because private tutoring for test preparation is such a pervasive feature of the education system, especially in Asia where high-stakes testing plays a prominent role in resource allocation in education. More generally, tutoring or buying disability designation resembles an “influence activity” as described by Milgrom (1988) and Milgrom and Roberts (1988). Much of this literature in organizational economics (e.g., Prendergast and Topel 1996; Ederer, Holden and Meyer 2018; Li, Mukherjee and Vasconcelos 2021) focuses on individual behavior: how individuals respond to incentives in unproductive ways and how to design contracts to reduce gaming. It will be interesting to study how informational externality among individuals and the sequential nature of the selection process affect organizational outcomes, as we do in this paper in the context of a selective admissions system.
Appendix

A. Proof of Proposition 5

The university chooses \((X_1, X_2, Y_1, Y_2) \in [0, 1]^4\) to maximize the mass of high-ability students admitted: \(V = q_1\lambda_1X_1 + (1 - q_1)\lambda_2X_2\), subject to the feasibility constraint. Because \(F(\cdot)\) satisfies monotone hazard rate, \((1 - F(s))\) is strictly concave on \((0, \hat{S})\), where \(\hat{S}\) is the (unique) maximizer of \((1 - F(s))\). It follows that \(\hat{S} \in (0, B)\). In the following, we let \(\theta(s) \equiv (1 - F(s))\) and \(\Theta(s) \equiv (\lambda + (1 - \lambda)F(s))\). Therefore, \(\theta(\cdot)\) is strictly concave on \((0, \hat{S})\) and linear on \((-\infty, 0)\), while \(\Theta(\cdot)\) is strictly convex on \((0, \hat{S})\) and linear on \((-\infty, 0)\).

Let \(s\) denote the cost type that is indifferent between tutoring and not in the first stage. Let \(C_i\) denote the cost type that is indifferent between tutoring and not in the second stage in School \(i\) \((i = 1, 2)\). With a slight abuse of notation, in this proof we allow \(s, C_2\) to be outside the range \([0, B]\) and allow \(C_1\) to be outside the range \([0, s]\), because they are defined to be indifferent types rather than the cutoff types. For example, if the indifferent type \(C_1 < 0\), then \(F(C_1) = 0\). As will be shown later, at the optimal policy the indifferent type and the cutoff type may only differ when the indifferent type is above the cutoff type in School 1.

We first show that a policy that results in no first-stage tutoring is not optimal.

Claim. Under the optimal policy, \(s > 0\).

Proof of claim. We establish this claim by showing that \(X_1 > X_2\) under the optimal policy, so the optimal policy will induce an indifferent type with positive cost to choose tutoring in the first stage.

Step 1. We show that among policies with \(X_1 \leq X_2\), the policy \(X_1 = X_2\) is optimal. When \(X_1 \leq X_2\), all low-ability students go to School 2, so we have \(G_2(C_2) = F(C_2)\) and \(\lambda_1 = 1\), the best policy with \(X_1 \leq X_2\) solves:

\[
\max_{X_1, X_2, C_2} \quad q_1X_1 + (\lambda - q_1)X_2 \\
\text{subject to} \quad q_1X_1 + (1 - q_1)X_2 - (1 - q_1)(1 - \lambda_2)(1 - F(C_2))\frac{C_2}{B} = Q, \\
X_2 \geq \frac{C_2}{B}, \quad X_1 \leq X_2.
\]

Suppose \(X_1 < X_2\) at the optimal solution, then it must be that \(X_2 = C_2/B\), because other-
wise, one can reduce $X_2$ by $q_1 \epsilon$ and increase $X_1$ by $(1 - q_1) \epsilon$ with positive $\epsilon$ small enough to improve the objective function value and keep the constraints all satisfied. Therefore, $X_1 < X_2$ implies $X_2 = C_2/B$. Then, $X_1 < C_2/B$ must be the solution to:

$$\max_{x_1,c_2} q_1 Bx_1 + (\lambda - q_1)c_2$$

subject to

$$q_1 Bx_1 + (1 - q_1)c_2 - (1 - q_1)(1 - \lambda_2)(1 - F(c_2))c_2 = Bq,$$

$$Bx_1 \leq c_2.$$  

This further simplifies to:

$$\max_{c_2} Bq - (1 - \lambda)F(c_2)c_2$$

subject to

$$(\lambda + (1 - \lambda)F(c_2))c_2 \geq 0.$$  

Because the objective function is decreasing in $c_2$, the constraint has to be binding, which contradicts $X_1 < C_2/B$.

Step 2. We next show that a policy with $X_1 = X_2$ is suboptimal. Because $X_1 = X_2 = C_2/B + Y_2$, we have $V = (\lambda/B)(C_2 + BY_2)$, where $C_2$ and $Y_2$ are subject to the feasibility constraint:

$$q_1 c_2 + (1 - q_1)(\lambda_2 + (1 - \lambda_2)G_2(c_2))c_2 + BY_2 = Bq.$$  

Since $G_2(c_2) = F(c_2)$ and $\lambda_1 = 1$, the feasibility constraint simplifies to:

$$C_2 + BY_2 - (1 - \lambda)(1 - F(c_2))c_2 = Bq.$$  

Let $V^*$ represent the value of the objective function when $C_1 = C_2 = s = s^*$. Then,

$$B(V^* - V) = \lambda[(1 + \delta)s^* - (C_2 + BY_2)]$$

$$= \lambda[(1 + \delta)s^* - Bq - (1 - \lambda)(1 - F(c_2))c_2]$$

$$= \lambda[(1 + \delta)s^* - (1 + \delta)s^*(\lambda + (1 - \lambda)F(s^*)) - (1 - \lambda)(1 - F(c_2))c_2]$$

$$= \lambda(1 - \lambda)[(1 + \delta)(1 - F(s^*))s^* - (1 - F(c_2))c_2].$$

(i) If $c_2 \leq s^* \leq \hat{s}$, then because $(1 - F(s))s$ is strictly increasing over $[0, \hat{s}]$, $(1 - F(s^*))s^* \leq (1 - F(c_2))c_2$, which implies $V^* > V$. (ii) If $s^* < c_2 \leq (1 + \delta)s^*$, it also implies $V^* > V$. (iii) If $(1 + \delta)s^* < c_2$, then $c_2 - (1 - \lambda)(1 - F(c_2))c_2 = (\lambda + (1 - \lambda)F(c_2))c_2 > (1 + \delta)(\lambda + (1 - \lambda)F(s^*))s^* = Bq$, which contradicts the feasibility constraint. So case (iii) is not
possible. The remaining possibility is (iv) \( s^* > \hat{s} \), in which case the equilibrium allocation has \( C_1 = C_2 = \hat{s} \). This will result in \( B(V^* - V) = \lambda(1 - \lambda)[(1 + \delta)\theta(\hat{s}) - \theta(C_2)] > 0 \). Therefore, in all cases the policy \( X_1 = X_2 \) is strictly dominated by the no-commitment equilibrium allocation, and hence cannot be optimal. \( \square \)

Let \( p \) denote the probability that a low-ability student with tutoring get admitted into School 1, which depends on the policy. If \( C_1 < C_2 \), the net benefit from first-stage tutoring for type \( c \) is

\[
\begin{cases} 
  p(BX_1 - BX_2) - \delta c & \text{if } c < C_1, \\
  p(BY_1 - (BX_2 - c)) - \delta c & \text{if } c \in [C_1, C_2], \\
  p(BY_1 - BY_2) - \delta c & \text{if } c > C_2.
\end{cases}
\]

The benefit is non-monotone in \( c \) if \( \delta < p \). In this case, we need to consider non-cutoff policies, where the set of types who obtains tutoring in Stage 1 may take the form \([0, s] \cup [s, \bar{s}]\). If \( C_1 \geq C_2 \) or \( \delta > p \), then all policies are cutoff policies, where types below \( s \) choose tutoring in Stage 1. We consider these policies in turn.

Cutoff policies. There are four possible types of cutoff policies: (1) \( C_1 \leq s \) and \( C_2 \geq s \); (2) \( C_1 \leq s \) and \( C_2 \leq s \); (3) \( C_1 \geq s \) and \( C_2 \geq s \); and (4) \( C_1 \geq s \) and \( C_2 \leq s \).

Consider case (1), \( C_1 \leq s \) and \( C_2 \geq s \), where we restrict attention to parameters such that \( \delta \geq p \) because we are dealing with cutoff policies.

The feasibility condition is:

\[
q_1(\lambda_1 + (1 - \lambda_1)G_1(C_1))X_1 + q_1(1 - \lambda_1)(1 - G_1(C_1))Y_1 + (1 - q_1)(\lambda_2 + (1 - \lambda_2)G_2(C_2))X_2 + (1 - q_2)(1 - \lambda_2)(1 - G_2(C_2))Y_2 = Q.
\]

Note that \( X_2 = Y_2 + C_2/B \) and \( X_1 = Y_1 + C_1/B \). The feasibility condition reduces to:

\[
q_1BY_1 + q_1(\lambda_1 + (1 - \lambda_1)G_1(C_1))C_1 + (1 - q_1)BY_2 + (1 - q_1)(\lambda_2 + (1 - \lambda_2)G_2(C_2))C_2 = BQ.
\]

The indifference condition in the first stage for type \( s \) is:

\[
\delta s = \frac{q_1}{\lambda + (1 - \lambda)F(s)}(BY_1 - BX_2 + s).
\]
Plugging in $X_2 = Y_2 + C_2/B$, the indifference condition gives:

$$q_1(BY_1 - BY_2) = \delta(\lambda + (1 - \lambda)F(s))s + q_1C_2 - q_1s.$$  

Incorporating this equation into the feasibility condition gives:

$$BY_2 + \delta(\lambda + (1 - \lambda)F(s))s + C_2 - q_1s + q_1(\lambda + (1 - \lambda)G_1(C_1))C_1 - (1 - q_1)(1 - \lambda)(1 - G_2(C_2))C_2 = BQ.$$  

Because $C_1 \leq s$ and the first stage has a cutoff nature, we have $G_1(C_1) = F(C_1)/F(s)$ and $G_2(C_2) = [F(C_2) - pF(s)]/[1 - pF(s)]$, where $p = q_1/(\lambda + (1 - \lambda)F(s))$ is the probability of admission into School 1. This implies

$$q_1(\lambda + (1 - \lambda)G_1(C_1)) = p(\lambda + (1 - \lambda)F(C_1)), \quad (1 - q_1)(1 - \lambda)(1 - G_2(C_2)) = (1 - \lambda)(1 - F(C_2)).$$

Plugging these two equations into the feasibility condition and simplifying, we have:

$$BY_2 + (\delta - p)(\lambda + (1 - \lambda)F(s))s + p(\lambda + (1 - \lambda)F(C_1))C_1 + (\lambda + (1 - \lambda)F(C_2))C_2 = BQ.$$  

Therefore, the constraint $Y_2 \geq 0$ is equivalent to:

$$BQ - (\delta - p)\Theta(s) - p\Theta(C_1) - \Theta(C_2) \geq 0.$$  

Similarly, plugging $X_i = Y_i + C_i/B$ ($i = 1, 2$) into the objective $V$ gives:

$$BV = \frac{\lambda}{\lambda + (1 - \lambda)F(s)}(q_1BY_1 - q_1BY_2 + q_1C_1 - q_1C_2) + \lambda BY_2 + \lambda C_2.$$  

Substituting out $q_1(BY_1 - BY_2)$ using the indifference condition obtained earlier, $BV$ becomes:

$$\frac{\lambda}{\lambda + (1 - \lambda)F(s)}[\delta(\lambda + (1 - \lambda)F(s))s - q_1s + q_1C_1] + \lambda BY_2 + \lambda C_2.$$  

This is equivalent to maximizing:

$$(\delta - p)s + pC_1 + BY_2 + C_2$$  

$= (\delta - p)s + pC_1 + C_2 + BQ - (\delta - p)(\lambda + (1 - \lambda)F(s))s - p(\lambda + (1 - \lambda)F(C_1))C_1 - (\lambda + (1 - \lambda)F(C_2))C_2$  

$= (1 - \lambda)[(\delta - p)(1 - F(s))s + p(1 - F(C_1))C_1 + (1 - F(C_2))C_2] + BQ.$
Therefore, the optimal policy in case (1) solves:

\[
\max_{s, C_1, C_2} (\delta - p)\theta(s) + p\theta(C_1) + \theta(C_2)
\]

subject to \(BQ - (\delta - p)\Theta(s) - p\Theta(C_1) - \Theta(C_2) \geq 0,\)
\[s - C_1 \geq 0, \quad C_2 - s \geq 0.\]

Consider a policy with \(s > \hat{s}.\) Lowering \(s\) weakly increases the objective function because \(\theta(\cdot)\) is decreasing for \(s > \hat{s}.\) Lowering \(s\) strictly relaxes the feasibility constraint because \(\Theta(\cdot)\) is increasing. Therefore, \(s > \hat{s}\) cannot be optimal. Similarly, \(C_i > \hat{s}\) cannot be optimal for \(i = 1, 2.\)

Consider any feasible policy that satisfies all the constraints with \((C_1, C_2, s) \in (-\infty, \hat{s}]^3\) not all equal. Then, there is an alternative policy \((C'_1, C'_2, s')\) such that

\[C'_1 = C'_2 = s' = \frac{\delta - p}{1 + \delta}s + p \frac{p}{1 + \delta}C_1 + \frac{1}{1 + \delta}C_2,\]

If the original policy satisfies the feasibility constraint, the alternative policy also satisfies the feasibility constraint because \(\Theta(\cdot)\) is convex. It strictly increases the objective function because \(\theta(\cdot)\) is strictly concave on \((0, \hat{s}],\) and \(C_2\) and \(s\) are both positive. It is therefore without loss of generality to only consider policies with \(C_1 = C_2 = s = \hat{s}.\) For such policies, the problem reduces further to:

\[
\max_{\hat{s} \in [0, \hat{s}]} (1 + \delta)\theta(\hat{s})
\]

subject to \(BQ - (1 + \delta)(s - (1 - \lambda)\theta(\hat{s})) \geq 0.\)

The solution to the problem is \(\hat{s} = \hat{S}\) if \(\hat{S} < s^*\) and \(\hat{s} = s^*\) otherwise. The policy corresponding to \(C_1 = C_2 = s = \hat{s}\) is \(Y_2 = Q - \frac{1 + \delta}{B}(\lambda + (1 - \lambda)F(\hat{s}))\hat{s},\) \(X_1 = Y_2 + \frac{\hat{s}}{B} + \frac{\delta(Q - Y_2)}{(1 + \delta)q},\) and \(X_2 = Y_2 + \frac{\hat{s}}{B}.\) Because no one gets a low score in School 1 under the optimal policy, the value of \(Y_1\) is not uniquely pinned down as long as all types in School 1 below \(\hat{s}\) weakly prefer to get tutoring at stage two. This requires \(Y_1 \in [0, X_1 - \frac{\hat{s}}{B}].\)

The analyses of the other three cases are similar to that of case (1), except that the weights attached to \(\theta(\cdot)\) (in the objective function) and \(\Theta(\cdot)\) (in the feasibility constraint) are different. For case (2), the objective function is \(\delta \theta(s) + p\theta(C_1) + (1 - p)\theta(C_2).\) In this case, it is optimal to set \(C_1 = C_2 = s.\) For case (3), the objective function is \(\theta(C_2) + \delta \theta(s).\)
In this case it is optimal to set \( C_2 = s \). For case (4), the objective function is \((1-p)\theta(C_2) + (\delta + p)\theta(s)\). In this case it is optimal to set \( C_2 = s \).

**Non-cutoff policies.** We consider next policies with the non-cutoff nature, with indifferent types \( \underline{s} < C_1 < s < C_2 < \bar{s} \). By their definitions,

\[
C_1 = B(X_1 - Y_1), \quad C_2 = B(X_2 - Y_2), \\
\delta_s = p(BX_1 - BX_2), \quad \delta s = p(BY_1 - BX_2 + s), \quad \delta \bar{s} = p(BY_1 - BY_2);
\]

where \( p = q_1/(\lambda + \lambda \bar{F}) \), with \( \bar{F} = F(s) + F(\bar{s}) - F(s) \).

Following the same steps as before, the feasibility condition can be written as:

\[
BY_2 + \delta(\lambda + (1-\lambda)\bar{F})s + C_2 - q_1 s + q_1(\lambda_1 + (1-\lambda_1)G_1(C_1))C_1 - (1-q_1)(1-\lambda_2)(1-G_2(C_2))C_2 = BQ.
\]

Note that

\[
q_1(\lambda_1 + (1-\lambda_1)G_1(C_1)) = p(\lambda + (1-\lambda)F(s)), \\
(1-q_1)(1-\lambda_2)(1-G_2(C_2)) = (1-\lambda)(1-pF(\bar{s}) - (1-p)F(C_2)), \\
G_2(C_2) = \frac{F(C_2) - pF(s) - p(F(C_2) - F(s))}{1-p\bar{F}},
\]

and \( G_1(C_1) = F(s)/\bar{F} \). Plugging these into the feasibility condition, we obtain:

\[
BY_2 - (p - \delta)(\lambda + (1-\lambda)\bar{F})s + p(\lambda + (1-\lambda)F(s))C_1 + p(\lambda + (1-\lambda)F(\bar{s}))C_2 + (1-p)(\lambda + (1-\lambda)F(C_2))C_2 = BQ.
\]

From the definition of \( \underline{s}, s, \bar{s}, C_1 \) and \( C_2 \), we have:

\[
C_1 = \frac{\delta}{p} \underline{s} + \frac{p - \delta}{p} s, \quad C_2 = \frac{\delta}{p} \bar{s} + \frac{p - \delta}{p} s.
\]

Substitute out \( C_1 \) and \( C_2 \) in all but the last term in the feasibility condition to get,

\[
BY_2 + \delta(\lambda + (1-\lambda)F(\underline{s}))\underline{s} + \delta(\lambda + (1-\lambda)F(\bar{s}))\bar{s} + (p - \delta)(\lambda + (1-\lambda)F(s))s + (1-p)(\lambda + (1-\lambda)F(C_2))C_2 = BQ.
\]

Therefore, the condition \( Y_2 \geq 0 \) is equivalent to:

\[
BQ - \delta \Theta(\underline{s}) - \delta \Theta(\bar{s}) - (p - \delta)\Theta(s) - (1-p)\Theta(C_2) \geq 0.
\]
Similarly, plugging the indifference conditions for $C_1$, $C_2$ and $s$ into the objective function gives:

$$BQ = \frac{\lambda}{\lambda + (1 - \lambda)\bar{F}} \left[ \delta(\lambda + (1 - \lambda)\bar{F})s - q_1s + q_1C_1 \right] + \lambda BY_2 + \lambda C_2.$$ 

This is equivalent to maximizing:

$$-(p - \delta)s + pC_1 + C_2 + BY_2 = BQ + (1 - \lambda)[-(p - \delta)(1 - \bar{F})s + p(1 - F(\bar{s}))C_1 + p(1 - F(s))C_2 + (1 - p)(1 - F(C_2))C_2].$$

Substituting out $C_1$ and $C_2$ in all but the last term from the above expression, the problem reduces to:

$$\max_{s, \bar{s}, C_2} \delta \theta(s) + \delta \theta(\bar{s}) + (p - \delta)\theta(s) + (1 - p)\theta(C_2)$$

subject to

$$C_2 = \frac{\delta}{p} \bar{s} + \frac{p - \delta}{p} s,$$

$$BQ - \delta \Theta(s) - \delta \Theta(\bar{s}) - (p - \delta)\Theta(s) - (1 - p)\Theta(C_2) \geq 0,$$

$$s - \bar{s} \geq 0, \quad \bar{s} - s \geq 0.$$

The same argument as before shows that we only need to consider policies with $(\bar{s}, s, \bar{s}, C_2) \in (-\infty, \hat{S}]^4$, where $\bar{s} > 0$. Furthermore, Jensen’s inequality implies that any such policy is dominated by a policy with $\bar{s} = s = \bar{s} = C_2$. So non-cutoff policies are never optimal.

**B. Proof of Proposition 6**

We first prove the counterpart to Lemma 1(c) to show that $C_1 = C_2$ for any $\omega \in [0, 1)$. Note that the support of $G_1$ is $[0, s]$ and $G_1(c) = F(c)/F(s)$ for $c \leq s$. For any such $c$, the accounting identity (2) is modified in this case to

$$q_1(1 - \lambda_1)G_1(c) + (1 - q_1)(1 - \lambda_2)G_2(c) = (1 - \omega)(1 - \lambda)F(c), \quad (10)$$

because a fraction $\omega$ of low-ability students who get first-stage tutoring have become high-ability students when they are in high school. Moreover, the fraction of high-ability students in School 1 is:

$$\lambda_1 = \frac{\lambda + (1 - \lambda)F(s)\omega}{\lambda + (1 - \lambda)F(s)}. \quad (11)$$
From the accounting identity (10), we obtain:

\[
\frac{G_2(C_1)}{G_1(C_1)} = 1 - \frac{(1 - \omega)(1 - \lambda)}{(1 - q_1)(1 - \lambda_2)} \left( 1 - \frac{F(C_1)}{G_1(C_1)} \right) \\
= \frac{1}{(1 - q_1)(1 - \lambda_2)} \left( (1 - \omega)(1 - \lambda) \frac{F(C_1)}{G_1(C_1)} - q_1(1 - \lambda_1) \right),
\]

where the second equality uses the relation, \( q_1(1 - \lambda_1) + (1 - q_1)(1 - \lambda_2) = (1 - \omega)(1 - \lambda) \). Hence,

\[
\frac{G_2(C_1)}{G_2(C_1)} \cdot \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2)} = \frac{1}{(1 - q_1)(1 - \lambda_2)} \left( (1 - \omega)(1 - \lambda) \frac{F(C_1)}{G_1(C_1)} - q_1(1 - \lambda_1) - \frac{(1 - q_1)\lambda_2(1 - \lambda_1)}{\lambda_1} \right) \\
= \frac{1}{(1 - q_1)(1 - \lambda_2)} \left( (1 - \omega)(1 - \lambda) \frac{F(C_1)}{G_1(C_1)} - \frac{(\lambda + (1 - \lambda)F(s)\omega)(1 - \lambda_1)}{\lambda_1} \right) \\
= \frac{(1 - \omega)(1 - \lambda)}{(1 - q_1)(1 - \lambda_2)} \left( \frac{F(C_1)}{G_1(C_1)} - F(s) \right),
\]

where the third equality uses (11) to substitute out \((1 - \lambda_1)/\lambda_1 \). Now, since \( G_1(C_1) = F(C_1)/F(s) \), the above expression is equal to 0. But recall that, for any \( \omega \in [0, 1) \), the equal credibility condition can be written as

\[
\frac{G_2(C_2)}{G_1(C_1)} = \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2)}.
\]

The previous two equations imply that \( C_1 = C_2 \). There are two possibilities: (1) \( C_1 = C_2 = s \); or (2) \( C_1 = C_2 < s \). We consider them in turn.

Case (1). If \( C_1 = C_2 = s \), then \( G_1(C_1) = 1 \). The university’s feasibility constraint becomes:

\[
q_1X_1 + (1 - q_1)(\lambda_2 + (1 - \lambda_2)G_2(s))X_2 = Q.
\]

When \( C_2 = s \), we have \( X_2(s) = s/B \). Furthermore, using the accounting identity (10) and the relation, \( q_1\lambda_1 + (1 - q_1)\lambda_2 = \lambda + (1 - \lambda)F(s)\omega \), the feasibility constraint reduces to:

\[
q_1X_1 + (\lambda + (1 - \lambda)F(s) - q_1)\frac{s}{B} = Q.
\]

From this we obtain

\[
X_1(s) = \frac{1}{q_1} \left( Q - (\lambda + (1 - \lambda)F(s) - q_1)\frac{s}{B} \right). \quad \text{(12)}
\]
Because cost-type $s$ will be indifferent between getting tutoring or not if she goes to School 2 (i.e., $BX_2 = s$), her benefit from first-stage tutoring is,

$$\frac{q_1}{\lambda + (1 - \lambda)F(s)}(\omega BX_1 + (1 - \omega)(BX_1 - s)) + \left(1 - \frac{q_1}{\lambda + (1 - \lambda)F(s)}\right)(\omega BX_2 + (1 - \omega)(BX_2 - s)) = \frac{q_1}{\lambda + (1 - \lambda)F(s)}B(X_1 - X_2) + \omega s.$$

If type-$s$ is indifferent, the benefit from first-stage tutoring is equal to the cost, $\delta s$. Substituting the values of $X_1$ and $X_2$ derived earlier, this indifference condition can be written as

$$(\lambda + (1 - \lambda)F(s))s = \frac{BQ}{1 + \delta - \omega}.$$

Let $s^*$ be the solution to the above equation. Then $C_1 = C_2 = s = s^*$ constitutes an equilibrium if the value of $X_1(s^*)$ specified in (12) satisfies $X_1(s^*) \geq X_2(s^*)$. The restriction that $X_1(s^*) \geq X_2(s^*)$ is equivalent to

$$BQ \geq (\lambda + (1 - \lambda)F(s^*))s^*.$$

This restriction would fail if $\omega > \delta$. Therefore, case (1) only obtains when $\omega \leq \delta$. Moreover, for $\omega \leq \delta$, the value of $s^*$ must be strictly less than $B$, and hence the indifference condition is indeed satisfied. Finally, it is obvious from the indifference condition that the equilibrium value of $s^*$ increases in $\omega$.

Case (2). $C_1 = C_2 < s$. Let $C$ represent the common value of $C_1$ and $C_2$. In this case, a type-$C$ student must be indifferent between tutoring or not in the second stage, regardless of which high school she attends. Therefore, we have $X_1 = X_2 = C/B$. The university’s feasibility constraint can be written as:

$$[q_1(\lambda_1 + (1 - \lambda_1)G_1(C)) + (1 - q_1)(\lambda_2 + (1 - \lambda_2)G_2(C))] \frac{C}{B} = Q.$$

This is equivalent to

$$[\lambda + (1 - \lambda)F(s)\omega + (1 - \omega)(1 - \lambda)F(C)] \frac{C}{B} = Q. \quad (13)$$

For case (2) to obtain, a type-$s$ student who chooses tutoring at the first stage expects to choose no tutoring at the second stage. Her benefit from first-stage tutoring comes from the possibility that she may become a high-ability student, in which case her probability
of admission will be $C/B$. Therefore, the indifference condition for type-$s$ requires

$$\delta s = \omega BX_1 = \omega C. \quad (14)$$

Of course we require $\omega > \delta$ for (14) to hold. An equilibrium in this case is a pair $(s^*, C^*)$ that satisfies equations (13) and (14) (if type $s^*$ is indifferent in the first-stage).

Equation (13) implicitly defines a function, $C = \gamma(s)$, with $\gamma(\cdot)$ decreasing and $\gamma(S^*) = S^*$ (where $S^*$ is the equilibrium cutoff for the one-stage model given by Proposition 1). For $s > S^*$, we have $C = \gamma(s) < s$. Therefore if $\gamma(B) < \delta B/\omega$, then there exists a unique solution $(s^*, C^*)$ to (13) and (14) with $s^* > C^*$. If $\gamma(B) > \delta B/\omega$, then the equilibrium cutoff is $s^* = B$ and $C^* = \gamma(B)$. In the latter case, equation (14) holds as an inequality. A student of type $s = B$ strictly prefers to obtain first-stage tutoring but will choose not to get tutoring in the second stage. Finally, it is obvious from equations (13) and (14) that $s^*$ increases in $\omega$. 

42
References


C. Supplementary Materials for Section 5.2

Proposition A1. In the two-stage setup with \( \pi \in (0, 1) \), an equilibrium exists. There is an indifferent low-ability cost-type \( s^* \in (0, B) \) such that low-ability students with types lower than or equal to \( s^* \) pay for tutoring in both stages, and high-ability students with types lower than or equal to \( s'^* = (1 - \pi)s^*/\pi \) pay for tutoring in both stages. No one else pays for tutoring. The equilibrium admissions policy of the university has \( X^*_1 > X^*_2 = s^*/B \), and \( Y^*_1 = Y^*_2 = 0 \).

Proof. Let \( G_i \) and \( G'_i \) represent the cost-type distributions in School \( i \) \( (i = 1, 2) \) for low-ability and high-ability students, respectively. The cost distributions in School 1 are:

\[
G_1(c) = \begin{cases} 
\frac{F(c)}{F(s)} & \text{for } c \in [0, s), \\
1 & \text{for } c \in [s, B];
\end{cases}
\]

\[
G'_1(c) = \begin{cases} 
\frac{F(c)}{F(s') + (1-F(s'))\pi} & \text{for } c \in [0, s'), \\
\frac{F(s') + (F(c)-F(s'))\pi}{F(s') + (1-F(s'))\pi} & \text{for } c \in [s', B].
\end{cases}
\]

The cost distributions in School 2 is determined by the identities:

\[
q_1(1-\lambda_1)G_1(c) + (1-q_1)(1-\lambda_2)G_2(c) = (1-\lambda)F(c),
\]

\[
q_1\lambda_1G'_1(c) + (1-q_1)\lambda_2G'_2(c) = \lambda F(c).
\]

Further note that first-stage selection implies that the average quality of School 1 is given by

\[
\lambda_1 = \frac{\lambda(F(s') + (1-F(s'))\pi)}{\lambda(F(s') + (1-F(s'))\pi) + (1-\lambda)F(s)\pi},
\]

with \( \lambda_2 = (\lambda - q_1\lambda_1)/(1-q_1) \).

In this model, the equal credibility condition can be written (in odds ratio form) as:

\[
\frac{\lambda_1 G'_1(C'_1) + (1-G'_1(C'_1))\pi}{1-\lambda_1 G_1(C_1)\pi} = \frac{\lambda_1 G'_2(C'_2) + (1-G'_2(C'_2))\pi}{1-\lambda_2 G_2(C_2)\pi},
\]
which is equivalent to:

\[
\frac{G_2(C_2)}{G_1(C_1)} \left( \frac{G_1'(C_1') + (1 - G_1'(C_1'))\pi}{G_2'(C_2') + (1 - G_2'(C_2'))\pi} \right) = \frac{\lambda_2(1 - \lambda_1)}{\lambda_1(1 - \lambda_2)}.
\] 

(A1)

**Claim.** If \( C_1 = C_2 = s \) and \( C_1' = C_2' = s' \), then the equal credibility condition is satisfied.

**Proof of claim.** Because \( C_1 = s \) and \( C_1' = s' \), we have \( G_1(C_1) = 1 \) and \( G_1'(C_1') = 1 \). The left-hand-side of equation (A1) is equal to

\[
\frac{\lambda_2}{1-\lambda_2} \frac{[(1 - \lambda)F(s) - q_1(1 - \lambda_1)]((1 - \pi)G'_1(s') + \pi) - (1 - \lambda - \lambda_1)q_1((1 - \pi)G'_1(s') + \pi)}{(1 - \pi)\lambda F(s') + \pi \lambda - \lambda_1 q_1((1 - \pi)G'_1(s') + \pi)}
\]

which is equal to the right-hand-side of (A1) because the definition of \( \lambda_1 \) and \( G'_1 \) imply:

\[
\frac{1 - \lambda_1}{\lambda_1} = \frac{(1 - \lambda)F(s)((1 - \pi)G'_1(s') + \pi)}{(1 - \pi)\lambda F(s') + \pi \lambda}.
\]

This establishes the claim. \( \square \)

We next derive the necessary conditions of an equilibrium satisfying \( C_1 = C_2 = s \) and \( C_1' = C_2' = s' \). Note that the benefit of second-stage tutoring for a low-ability student is less than \( \pi B \). Hence we must have \( C_1 = C_2 = s < B \). Therefore, there is an indifferent cost-type among low-ability students in the first-stage, whose indifference condition is:

\[
\delta s = \pi B \frac{X_1 - X_2}{\lambda(\pi F(s') + (1 - \pi F(s'))\pi) + (1 - \lambda)F(s)\pi}.
\]

Because the benefit of first-stage tutoring of a high-ability student is \( (1 - \pi)/\pi \) times the similar benefit for a low-ability student, the cutoff cost-type among the high-ability students is \( s' = (1 - \pi)s/\pi \). Substituting out \( s' \) using \( s' = (1 - \pi)s/\pi \), \( \lambda_2 \), \( G_2 \) and \( G'_2 \) can all be written as functions of \( s \). Then \( X_1(s) \) is implied by the university’s feasibility condition:

\[
q_1X_1 + (1 - q_1)(\lambda_2(G_2'(s') + (1 - G_2'(s'))\pi) + (1 - \lambda_2)G_2(s)\pi\frac{s}{B} = Q.
\]

Note that \( X_1(s) \) thus defined is strictly decreasing in \( s \). Then the equilibrium \( s^* \) is defined
by the first-stage indifference condition:
\[ \delta^{s*} = \pi(BX_1(s^*) - s^*) \frac{q_1}{\lambda (F\left(\frac{1 - \pi}{\pi} s^*\right) + (1 - F\left(\frac{1 - \pi}{\pi} s^*\right)) \pi) + (1 - \lambda)F(s^*)\pi}. \]

The solution \( s^* \) is unique and positive. \( \blacksquare \)

D. Supplementary Materials for Section 5.3

We develop the model described in Section 5.3 in this appendix.

Let student ability be \( a \in \{a_n\}_{n=1}^{N} \) with \( a_n < a_{n'} \) for any \( n < n' \). Normalize the total mass of students to be 1. The mass of students with ability \( a_n \) is \( \lambda_n \). A higher ability student tends to get a higher score. A student with ability \( a_n \) obtains test core \( T = a_n + U \) if she does not get tutoring, where \( U \) is a continuous random variable with distribution \( \Phi \) on \(( -\infty, \infty)\). We assume that the corresponding density function \( \phi \) is strictly log-concave. This assumption ensures that, for \( T' > T \), the likelihood ratio, \( \phi(T' - a)/\phi(T - a) \) strictly increases in \( T \).

Tutoring allows \( a_n \) to mimic the next higher type \( a_{n+1} \) for \( n < N \). That is, type \( a_n \) gets score \( T = a_n + U \) with tutoring. Tutoring cost \( c \) is distributed according to distribution \( F_n \), which has a continuous and everywhere positive density function on \([0, B]\). We allow the cost distribution \( F_n \) to depend on type in an arbitrary way.

Let \( S^n \) denote the cutoff cost level such that students with cost \( c \leq S^n \) and ability \( a = a_n \) get tutoring. The likelihood ratio of obtaining test score \( T \) for type \( a_n \) relative to type \( a_{n-1} \) for \( n < N \) is:
\[ \frac{(1 - F_n(S^n))\phi(T - a_n) + F_n(S^n)\phi(T - a_{n+1})}{(1 - F_{n-1}(S^{n-1}))\phi(T - a_{n-1}) + F_{n-1}(S^{n-1})\phi(T - a_n)}. \]

This likelihood ratio is increasing in \( T \) because both \( \phi(T - a_n)/\phi(T - a_{n-1}) \) and \( \phi(T - a_{n+1})/\phi(T - a_n) \) increase in \( T \). Thus, regardless of the tutoring behaviors, the posterior belief about student ability is strictly increasing (in the sense of likelihood-ratio dominance) in the test score. The optimal admissions policy for a university that wants to maximize the average ability of its intake is a cutoff rule—a student is admitted if and only if her test score exceeds some cutoff \( \hat{T} \). For a student with ability \( a_n \), the benefit from tutoring is:
\[ \beta(a_n, \hat{T}) = B\left(\Phi(\hat{T} - a_n) - \Phi(\hat{T} - a_{n+1})\right). \]

By log-concavity of \( \phi \), \( \beta(a_n, \cdot) \) is first increasing and then decreasing in \( \hat{T} \) for any \( a_n \). For
sufficiently large $\hat{T}$, the partial derivative with respect to $\hat{T}$ is negative.

For the cutoff rule $\hat{T}$ to satisfy the university’s quota constraint, we require the mass of students with scores above the cutoff to be equal to the quota:

$$\sum_n \left[ F_n(\beta(a_n, \hat{T})) (1 - \Phi(\hat{T} - a_{n+1})) + (1 - F_n(\beta(a_n, \hat{T}))) (1 - \Phi(\hat{T} - a_n)) \right] \lambda_n = Q,$$

where we use the equilibrium restriction that $S^n = \beta(a_n, \hat{T})$. The left-hand-side of the above equation goes to 1 as $\hat{T} \to -\infty$ and goes to 0 as $\hat{T} \to \infty$. Therefore, there exists $\hat{T}$ that satisfies the equation. Consider the largest such solution among all solutions and denote it by $\hat{T}^*$. At such $\hat{T}^*$, the left-hand-side of the quota constraint crosses $Q$ from above, and so we have $\partial \hat{T}^*/\partial Q < 0$. Moreover, $\hat{T}^*$ approaches infinity as $Q$ approaches 0.

The equilibrium measure of students getting tutoring is:

$$\sum_n F_n(\beta(a_n, \hat{T}^*)) \lambda_n.$$

Because $\hat{T}^*$ is decreasing in $Q$ and $\beta(a_n, \cdot)$ is decreasing in $\hat{T}^*$ for $Q$ sufficiently small, the measure of students getting tutoring is increasing in $Q$ when competition for college admission is intense. Therefore, we have the counterpart to Corollary 1: in the largest equilibrium, the total amount of tutoring increases as the number of university places increases.

For the analysis of two stages of gaming, we restrict the general model to the case of $N = 2$ and let $a_L = a_1$, $a_H = a_2$. The proportion of high-ability students in School $i$ ($i = 1, 2$) is denoted $\lambda_i$. We use $\ell(T) = \phi(T - a_H)/\phi(T - a_L)$ to denote the likelihood ratio. We use $T^0$ to denote the test score that satisfies $\ell(T^0) = 1$.

For the first stage of selection, let test scores in the high-school entrance examination be given by $t = a_j + u$ for $j = L, H$, with $u$ being a random variable distributed according to $\Psi$, with a log-concave density $\psi$. If a low-ability students of cost type $c$ pays to get first-stage tutoring at cost $\delta c$, her test score will be $t = a_H + u$. The cost distribution among all low-ability students is $F$ on $[0, B]$. We continue to use $s$ to represent the first-stage cutoff type. If the admission threshold for School 1 in the high-school entrance examination is $\hat{t}$, we have

$$G_1(c) = \begin{cases} \frac{F(c)(1 - \Psi(\hat{t} - a_H))}{F(\hat{t})(1 - \Psi(\hat{t} - a_H)) + (1 - F(\hat{t}))(1 - \Psi(\hat{t} - a_L))} & \text{if } c \in [0, s), \\
\frac{F(c)(1 - \Psi(\hat{t} - a_H)) + (F(c) - F(\hat{t})) (1 - \Psi(\hat{t} - a_L))}{F(\hat{t})(1 - \Psi(\hat{t} - a_H)) + (1 - F(\hat{t}))(1 - \Psi(\hat{t} - a_L))} & \text{if } c \in [s, B]. \end{cases}$$

iv
Observe that \( G_1(\cdot) \) first-order stochastically dominates \( \max\{F(\cdot)/F(s), 1\} \).

The university’s admissions policy can be summarized by two test score thresholds, \((T_1, T_2)\), for students from Schools 1 and 2. Under this admissions policy, the feasibility condition of the university is:

\[
\sum_{i=1,2} q_i \left[ (\lambda_i + (1 - \lambda_i) G_i(C_i)) (1 - \Phi(T_i - a_H)) + (1 - \lambda_i) (1 - G_i(C_i)) (1 - \Phi(T_i - a_L)) \right] = Q.
\]

(A2)

The equal credibility condition require \( K_1(T_1) = K_2(T_2) \), which can be expressed (in odds ratio form) as:

\[
\frac{\lambda_1}{1 - \lambda_1} \frac{\ell(T_1)}{G_1(C_1) \ell(T_1) + 1 - G_1(C_1)} = \frac{\lambda_2}{1 - \lambda_2} \frac{\ell(T_2)}{G_2(C_2) \ell(T_2) + 1 - G_2(C_2)},
\]

(A3)

**Proposition A2.** In any second-stage subgame equilibrium with \( \lambda_1 > \lambda_2 \), the admissions standard for the selective school is strictly lower, i.e., \( T_1 < T_2 \).

**Proof.** Suppose to the contrary that \( T_1 \geq T_2 \). There are three cases.

(a) If \( T_1 \geq T^0 \geq T_2 \), then \( K_1(T_1) \geq \lambda_1 > \lambda_2 \geq K_2(T_2) \), violating the equal credibility condition.

(b) If \( T^0 \geq T_1 \geq T_2 \), then \( 1 \geq \ell(T_1) \geq \ell(T_2) \). Because \( \beta(a_L, \cdot) \) is increasing for such values of \( T \), we have \( \beta(a_L, T_1) \geq \beta(a_L, T_2) \). The indifference conditions at the two schools then imply \( C_1 \geq C_2 \). Furthermore, since School 1 tends to select students with low tutoring costs, \( G_2 \) first-order stochastically dominates \( G_1 \), and therefore \( G_1(C_1) \geq G_1(C_2) \geq G_2(C_2) \). Together, these inequalities imply

\[
\frac{\ell(T_1)}{G_1(C_1) \ell(T_1) + 1 - G_1(C_1)} \geq \frac{\ell(T_2)}{G_2(C_2) \ell(T_2) + 1 - G_2(C_2)},
\]

which contradicts (A3) whenever \( \lambda_1 > \lambda_2 \).

(c) If \( T_1 \geq T_2 \geq T^0 \), then we have \( \ell(T_1) > \ell(T_2) \). Recall that we define \( \rho = \lambda_1(1 - \lambda_2)/(\lambda_2(1 - \lambda_1)) \). The equal credibility condition (A3) implies

\[
\rho = \frac{\ell(T_2)}{G_2(C_2) \ell(T_2) + 1 - G_2(C_2)} \frac{G_1(C_1) \ell(T_1) + 1 - G_1(C_1)}{\ell(T_1)} < \frac{G_1(C_1) \ell(T_1) + 1 - G_1(C_1)}{G_2(C_2) \ell(T_1) + 1 - G_2(C_2)}.
\]
Since \( \ell(T_1) \geq 1 \) and \( \rho > 1 \), the above inequality implies \( G_1(C_1) > G_2(C_2) \). Therefore,

\[
\rho < \frac{G_1(C_1) \ell(T_1) + 1 - G_1(C_1)}{G_2(C_2) \ell(T_1) + 1 - G_2(C_2)} < \frac{G_1(C_1)}{G_2(C_2)}.
\]

For \( T_i \geq T^0 \), \( \beta(a_L, \cdot) \) is decreasing in the relevant region. So \( T_1 \geq T_2 \) implies \( \beta(a_L, T_1) \leq \beta(a_L, T_2) \), which in turn implies \( C_1 \leq C_2 \). We therefore obtain \( \rho < G_1(C_1)/G_2(C_1) \). However, we have \( G_i(C_i) \leq F(C_i)/F(s) \) due to first-order stochastic dominance. By the proof of Lemma 1(c), this is equivalent to \( \rho \geq G_1(C_1)/G_2(C_1) \), which leads to a contradiction. ■

Proposition A2 is premised on the condition that \( \lambda_1 > \lambda_2 \). Despite the fact that the two high schools have identical intrinsic quality, the following proposition shows that there exists an equilibrium in which endogenous differences between the two high schools can arise, so that \( \lambda_1 > \lambda_2 \) indeed arises as an equilibrium outcome, as gaming for selective admission unravels to the high-school entrance stage. The condition stated in Proposition A3 below means that if the university recruits exclusively from School 1, its standard must be higher than \( T^0 \) even if no student in that school chooses tutoring. This condition implies that any feasible admissions standard must satisfy \( T_1 > T^0 \), and by Proposition A2 we must have \( T_2 \geq T_1 > T^0 \) in equilibrium. It serves a similar purpose as Assumption 1 does for the binary score model, namely to ensure that strategic substitution prevails in the relevant region.

**Proposition A3.** If \( Q < q_1 [\lambda (1 - \Phi(T^0 - a_H)) + (1 - \lambda)(1 - \Phi(T^0 - a_L))] \), then there exists an equilibrium with \( T^*_2 > T^*_1 > T^0 \).

**Proof.** Fix a first-stage cutoff \( s \in [0, B] \). For such \( s \), there is a unique admissions standard \( t_0 \) that will fill the quota for School 1, given by:

\[
(\lambda + (1 - \lambda)F(s))(1 - \Psi(t_0 - a_H)) + (1 - \lambda)(1 - F(s))(1 - \Psi(t_0 - a_L)) = q_1.
\]

This gives \( \lambda_1 = \lambda(1 - \Psi(t_0 - a_H))/q_1 \) and \( \lambda_2 = (\lambda - q_1 \lambda_1)/(1 - q_1) < \lambda_1 \). For such \( \lambda_1, \lambda_2, \) and \( s \), the second-stage subgame must satisfy (A2), (A3), and the indifference conditions \( C_i = \beta(a_L, T_i) \) for \( i = 1, 2 \).

Consider a university policy \( (T_1, T_2) \in [T^0, 1]^2 \). The indifference conditions determine \( C_i \) as a decreasing function of \( T_i \). As \( T_1 \) increases, \( C_1 \) falls, and the university will admit fewer students from School 1. To satisfy the quota constraint (A2), it must lower \( T_2 \) (which will induce a corresponding increase in \( C_2 \)). Equation (A2) therefore defines an implicit
function, \( T_2 = \tau_{\text{fea}}(T_1) \), which is downward-sloping. At \( T_1 \), if the university quota would be exceeded even when \( T_2 = \infty \), we define \( \tau_{\text{fea}}(T_1) \) to be equal to \( \infty \). The assumption stated in the proposition implies \( \tau_{\text{fea}}(T^0) = \infty \). Obviously if \( T_1 \) goes to infinity, \( T_2 \) must be finite because no student would be admitted to university otherwise. We thus have \( \lim_{T_1 \to \infty} \tau_{\text{fea}}(T_1) < \infty \).

As \( T_1 \) increases and \( C_1 \) falls, the marginally admitted student from School 1 becomes more credible. To maintain equal credibility across the two schools requires \( T_2 \) to rise (which will induce \( C_2 \) to fall). Equation (A3) therefore defines an implicit function, \( T_2 = \tau_{\text{cre}}(T_1) \), which is upward-sloping, with \( \tau_{\text{cre}}(T^0) < \infty \) and \( \lim_{T_1 \to \infty} \tau_{\text{cre}}(T_1) = \infty \). This establishes that, for each \( s \), there exists a unique pair \( (T_1, T_2) \) such that \( T_2 = \tau_{\text{cre}}(T_1) \) and \( T_2 = \tau_{\text{feas}}(T_1) \). By Proposition A2, \( T_2 > T_1 > T^0 \).

For each \( s \), the corresponding standard \( \hat{\ell} \) for selection into School 1 is determined by:

\[
(\lambda + (1-\lambda)F(s))(1-\Psi(\hat{\ell} - a_H)) + (1-\lambda)(1-F(s))(1-\Psi(\hat{\ell} - a_L)) = q_1.
\]

Define \( \sigma(s) \) to be equal to:

\[
\min \left\{ \frac{1}{\delta} (\Psi(\hat{\ell} - a_L) - \Psi(\hat{\ell} - a_H)) [B(\Phi(T_2 - a_L) - \Phi(T_1 - a_L)) + \max\{\beta_L(a_L, T_1) - s, 0\}], B \right\},
\]

where \( \hat{\ell}, T_1 \) and \( T_2 \) are determined by \( s \). By investing in first-stage tutoring, a low-ability raises her chance of being admitted into School 1 from \( 1-\Psi(\hat{\ell} - a_L) \) to \( 1-\Psi(\hat{\ell} - a_H) \). The rent from being admitted to School 1 is given by the term is square brackets and is positive. Therefore, \( \sigma(\cdot) \) is a continuous mapping from \([0, B]\) to itself. There exists \( s^* \in [0, B] \) such that \( s^* = \sigma(s^*) \). The admissions standards \( \hat{\ell}^*, T_1^* \) and \( T_2^* \) corresponding to such cutoff \( s^* \), together with the associated second-stage cutoffs \( C_1^* \) and \( C_2^* \) determined by the indifference conditions, constitute an equilibrium of the model.