Do Peer Preferences Matter in School Choice Market Design? Theory and Evidence[*]

Natalie Cox[†], Ricardo Fonseca[‡] and Bobak Pakzad-Hurson[§]

July 26, 2021

Abstract

Can a clearinghouse generate a stable matching if it does not allow students to express their preferences over both programs and peers? Application data from Australia's centralized college admissions system show that students have preferences over the academic abilities of their peers. However, the matching mechanism used only allows students to express preferences over programs, not over peers. Theoretically, we show that a stable matching exists with peer preferences under mild conditions, but finding one via canonical mechanisms is unlikely. We show that increasing transparency about the previous cohort of students enrolling at each program, analogously to the process in the Australian market, induces a tâtonnement wherein the distributions of former students play the role of prices. We theoretically model this process and develop a test for match stability. We implement this test empirically to show that the Australian market fails to converge to stability over time, and that this instability especially affects low socioeconomic status students. To address these issues, we propose a new mechanism that improves upon the current design, and we show that this mechanism generates a stable matching in the Australian market.

[*] For very helpful comments and discussions we thank Eduardo Azevedo, Fuhito Kojima, Jacob Leshno, Margaux Luflade, George Mailath, Matt Pecenco, Rakesh Vohra, and seminar audience members. We are grateful to Joanna Tasmin and especially Clemens Lehner for excellent research assistance.

[†] Princeton University, Bendheim Center for Finance, 20 Washington Rd, Princeton, NJ 08540. Email: nbachas@princeton.edu.

[‡] Brown University, 8 Fones Alley, Providence, RI 02912. Email: ricardo_fonseca@brown.edu.

[§] Brown University, 64 Waterman Street, Providence, RI 02912. Email: bph@brown.edu

I Introduction

Creating a stable matching – a matching in which no individual wants to leave her partner(s) and rematch with another willing partner (or remain unmatched) – is often viewed as the chief concern in many market design settings (Roth, 2002). Following the application of matching theory to education markets (Abdulkadiroğlu and Sönmez, 2003), at least 46 countries now use centralized mechanisms to assign students to colleges (Neilson, 2019). The mechanisms used universally assume that there are no complementarities in student preferences; that is, students do not have preferences over their peers. However, a large literature has established the importance of peer *effects* on educational outcomes, and suggests the possibility of peer *preferences* at the college level.[1] Given this evidence, what happens if the matching mechanism is *misspecified*, in that students are only allowed to express preferences over college programs, but have preferences over both programs and their peers?

In this paper, we seek to answer three questions: Do students have peer preferences? Does a stable matching exist when students have peer preferences? What are the consequences of failing to account for peer preferences in a centralized matching mechanism? We study these questions theoretically and empirically, using data from Australia's centralized matching market for college admissions.

Theoretically, we find that a stable matching exists when students have peer preferences, under mild conditions. However, mechanisms used in practice that do not solicit student preferences over peers are unlikely to yield a stable matching. Guided by a common convention in real-world markets, we study a dynamic process in which students update their beliefs on their potential peers at each program using information from the previous cohort's matching. This induces a pseudo-tâtonnement process, and we derive a simple test for convergence to a stable matching. This process is not guaranteed to converge to a stable matching–in fact, it can possibly converge only for subsets of agents in the market–and we discuss sufficient conditions under which it converges to a stable matching in finite time.

Empirically, we use data from Australia's college admissions market to establish the existence of peer preferences and to test the stability of matchings generated by the existing assignment

[1] See Sacerdote (2011) for a literature review. At the college level, there is evidence of peer effects in college student achievement. Stinebrickner and Stinebrickner (2006) use survey data to find that roommates have an effect on student achievement, while Conley et al. (2018) find similar results using the study times of individuals in a social network. At the primary and secondary school levels, a series of recent papers show that a student's ordinal "ability" ranking within her school and class has a negative effect on educational achievement; that is, students perform worse when they have higher achieving peers (see Attewell (2001); Abdulkadiroğlu, Angrist, and Pathak (2014); Dobbie and Fryer Jr. (2014); Elsner and Isphording (2017); Elsner, Isphording, and Zölitz (2018); Murphy and Weinhardt (2020); Yu (2020); Zárate (2019); Carrasco-Novoa, Diez-Amigo, and Takayama (2021)). Abdulkadiroğlu, Angrist, and Pathak (2014) do not find a large effect of peer ability on student performance.

process. We observe how students' rankings over university programs respond to information about their own quality relative to that of their prospective peers at each program. While some recent research has studied educational settings in which agents prefer being matched to higher ability peers,[2] our data suggest that students prefer not to match with a program where they are near the bottom of the ability distribution. This pattern is in accordance with the "big-fish-little-pond effect," which has been well documented in the education literature.[3] Peer achievement is negatively correlated with a student's "self-concept," which particularly affects students at the bottom of the "ability" distribution (Pop-Eleches and Urquiola, 2013). Using data on program selectivity over more than a decade, we also test our theoretical predictions of convergence to a stable matching, and verify that the top "quality" part of the market converges while the bottom does not.

To begin our formal analysis, and to provide a foundation for our empirical findings, we construct a matching model with a continuum of students and finitely many programs, as in Azevedo and Leshno (2016). We depart from this model by assuming that student preferences depend on their intrinsic values over programs and the distribution of student abilities at each program, similarly to Leshno (2021). This can encompass cases where, for example, students wish to attend programs enrolling the highest-ability peers, or the opposite, where students wish to avoid more able colleagues.

As in an equilibrium of a club good economy (see e.g Ellickson et al. (1999) and Scotchmer and Shannon (2015)), a stable matching is endogenously supported by the set of students at each program; stability requires that no student wishes to block the matching in favor of another program that is willing to take her, *given the students already assigned to each program*. Under mild assumptions, a stable matching always exists. Unlike in Azevedo and Leshno (2016), the set of stable matchings is not generally a singleton.

How can a market designer ensure that a stable matching is created? Soliciting student preferences as functions of the sets of students attending each program may be both too complex for students to report (Zhang and Levin, 2017; Budish and Kessler, 2020) and outside the realm of consideration for many centralized clearinghouses (for a thorough discussion of this point, see Carroll (2018)). Canonical, static mechanisms–such at the celebrated deferred acceptance mechanism of Gale and Shapley (1962)–in which students are only able to list ordinal preferences over programs, and not over peers, may fail to deliver a stable matching. If the distribution of student preferences is common knowledge, the set of Nash equilibria of the games induced by "well-behaved" mechanisms coincides with the set of stable matchings; students are able to "roll in"

[2] See Rothstein (2006), Abdulkadiroğlu et al. (2020), Allende (2020), and Beuermann et al. (2019).

[3] See Marsh et al. (2008); Seaton, Marsh, and Craven (2009).

peer preferences to their reports to the mechanism. However, we show that when students do not have accurate beliefs of the preferences of other students, these mechanisms likely fail to generate a stable matching. As a key lesson of the market design literature is to avoid assumptions of common knowledge and sophistication – the so-called Wilson Doctrine (Wilson, 1987), we view this as a negative result.

Therefore, we instead focus our attention on how students' beliefs about their peers arise. We study a discrete-time dynamic process which mirrors that of our empirical setting: students observe the distribution of abilities enrolling at each program in the previous cohort, and submit a rank order list (ROL) over programs to a centralized matchmaker who delivers a stable matching with respect to the reported preferences in each period. This type of belief updating is common in many higher-education markets; for example, U.S. News and World Report publishes a popular annual publication, revealing test scores of entering classes from the previous year at US universities, as an aid to current applicants. That the matchmaker (perhaps naively) generates a stable matching with respect to reported rankings over programs reflects that peer preferences are not being explicitly considered in the matching process.

Under the assumption that fundamentals of the market do not change over time, and that students report their ordinal preferences for programs under the belief that the distribution in the current period will mirror that of the previous period, this market forms a discrete-time tâtonnement process. The distribution of student scores serves the role of prices in a typical exchange economy, and students best respond to the previous period's "prices," just as in the original Cournot updating procedure.[4] Unlike traditional tâtonnement processes, a matching is constructed in each period. Therefore, we refer to this as the *Tâtonnement with Intermediate Matching (TIM)* process.

Our main theoretical result provides a simple tool for an observer to judge the stability of a sequence of matchings in the TIM process: the distribution of student abilities at each program are (approximately) in steady state if and only if the market creates a (approximately) stable matching.

We identify three shortcomings of the TIM process in ensuring stability. First, it need not converge, even when there is a unique stable matching. In these cases, it will fail to generate a stable matching, even in the long run. Second, even if it does converge, it need not do so immediately, and therefore, instability persists along the path to stability. Third, the process is fragile to changes in the market; if there is, for example, entry and exit of programs between time periods, then little information may be transmitted across cohorts.

We suggest an alternative mechanism which more explicitly accounts for peer preferences, and improves upon the three problems we identify in the TIM process. This mechanism impor-

[4] This is also similar to the notion of fictitious play, proposed by Brown (1951). As Berger (2007) remarks, the simultaneous decisions made within cohort are actually a variant of the original fictitious play framework.

tantly differs in that it induces a tâtonnement procedure *within* each period's cohort of students, therefore, we call this the *Tâtonnement with Final Matching (TFM)* mechanism. Each cohort is broken up into smaller, randomly created sub-cohorts, and each sub-cohort submits their ordinal preferences over programs to the mechanism. These preferences are used to construct a "pseudo-matching" using a stable mechanism that is not consummated, but is used to provide information to the next sub-cohort; each sub-cohort is informed of the distribution of student abilities *only from the previous sub-cohort* at each program. This iteration continues until the distribution of student abilities changes by less than some prespecified amount between subsequent cohorts. At this point, all students, including those in sub-cohorts who have already submitted their preferences, are asked to submit their final preference lists to the matchmaker, who creates the final matching.

This iterative mechanism resembles those in use in higher-education markets in China, Brazil, Germany and Tunisia (see Bo and Hakimov (2019); Luflade (2019)), but importantly requires different sub-cohorts of students to report their preferences in a particular order so that the information of their pseudo-matching can be passed along to subsequent students.

This iterative TFM mechanism has several desirable properties. First, it generates a (approximately) stable matching whenever the TIM process converges to a stable matching in the long run, and does so without necessitating a string of unstable matchings for early cohorts. Second, for an appropriate starting condition, it generates a stable matching in a wide class of markets (i.e. in those markets where we can guarantee existence of a stable matching). This stands in contrast to the TIM mechanism. Third, it is not susceptible to instability caused by changes in market primitives (such as changes in underlying preferences or changes in the set of programs) as the mechanism does not rely on information from the previous cohorts. Fourth, it induces a game in which truthfully reporting one's ordinal preferences, taking the previous sub-cohort's pseudo-matching into account, is an $\epsilon-$Nash equilibrium. We show that these benefits come at little administrative cost for students; by appropriately selecting the number of sub-cohorts, only a small fraction of students will have to re-state their preferences.

Other papers have studied peer preferences in a centralized matching framework (Echenique and Yenmez, 2007; Bykhovskaya, 2020; Pycia, 2012; Pycia and Yenmez, 2019). These papers investigate small sets of students (i.e. the set of students is not a continuum, nor do the results look at limiting cases of many students) and are primarily concerned with conditions under which stable matchings exist and can be found. This contrasts to our setting, where a stable matching exists under very mild conditions. More similar to our paper is contemporaneous work by Leshno (2021). Indeed, our models are similar as they build on Azevedo and Leshno (2016). One main difference is that our model allows for students to care about the entire distribution of peer abilities, whereas Leshno's assumes students care only about summary statistics of student abilities. We provide theoretical results for this case in Section II.D. The focuses of our papers are also different, with Leshno providing several results on how the continuum model is a valid approximation of large, finite models. As a result, we do not pursue such findings in our similar setting.

We investigate the presence and impact of peer preferences using data from Australia's centralized market for college admissions. Australia matches students to programs using the deferred acceptance algorithm. Students are ranked predominantly based on the results of a standardized test, the ATAR. Importantly, in our setting, when applying to programs students have information on the ATAR scores of the cohort admitted to each program in the *previous* year. We refer to this going forward as the previous year statistic (PYS).

The ATAR score is a proxy for student ability as it predicts student academic performance at the university level (UAC study). As in our model, the PYS provides applicants with information on the ATAR score distribution, and hence ability distribution, of students admitted to each program in the previous year. There is anecdotal evidence that students base their program selection on the ATAR scores of the previous cohort. As one student says, "I was contemplating changing from Commerce/Engineering to Science/Engineering as other people who obtained a similar [ATAR] to myself were doing the Science double. Not many opt for commerce" (James, Baldwin, and McInnis, 1999).

In our dataset, we observe the universe of applicant ATAR scores, applicants' rank ordered lists, and program PYSs in New South Wales, the largest state in Australia from 2005 to 2018. We find evidence that students have peer preferences. More specifically, they have ordinal ranking concerns (Frank, 1985) wherein they prefer not to attend programs that admit peers with ATAR scores systematically above their own. The utility effect we infer from entering a program is asymmetric, and similar to the analogous function in Card et al. (2012); students face a utility loss only if their score is below the PYS, and the utility loss is increasing in the difference. We call these "big fish" preferences, in reference to the big-fish-little-pond analogy.

We use the data to establish the existence of these preferences in two ways. First, we look across time at the response of applicants to changes in programs' PYSs. In line with big fish preferences, as a program's PYS increases, it attracts applications from students with higher ATAR scores. Moreover, the response is asymmetric; students with scores below the PYS are less likely to rank the program.[5]

Second, we look *within* the same applicant over time, as they change their ROL before and after learning their ATAR scores.[6] We observe one snapshot of each applicant's ROL immediately

[5] We rule out the possibility that this effect is driven by trends in program quality over time, or by students learning about their own "fit" at a program, by including lagged PYSs and including program age fixed effects. The age fixed effects proxy for the amount of ambient information students have about programs unrelated to peer effects.

[6] Narita (2018) uses a similar identification strategy in a centralized market for high school admission.

before they learn their own ATAR score, and one after. Big fish preferences predict than an applicant will adjust their ROL to prioritize schools with similar PYSs after learning their true ATAR score.[7] Thus, we compare how students with initially similar ROLs respond to differential ATAR score results. After learning their ATAR scores, students alter their ROLs in the following way: they systematically *drop* programs with PYSs far above their own ATAR score, *add* new programs with PYSs closer to their own ATAR score, and *promote* the ranking of programs that were on their initial ROL and have PYSs closer to their own ATAR score.

While our paper is the first that we know of that argues for the existence of peer preferences at the university level, other papers show that primary and secondary school students (or, rather, their parents) have preferences over the other students attending a school. Allende (2020) incorporates peer preferences into a model of demand for primary education in Peru, and empirically documents a taste for high SES peers. Rothstein (2006), Abdulkadiroğlu et al. (2020), and Beuermann et al. (2019) find that students prefer schools with higher-achieving students. That university students in our setting avoid higher-scoring peers, and secondary school students in these papers seek out higher-scoring peers is an interesting difference. Beuermann and Jackson (2020) argue that parents have preferences over peers for their children that do not improve secondary-school exam performance, raising the possibility that students and parents have different utility functions.

We investigate the impact of peer preferences on stability in two markets when beliefs are governed by the distribution of student types enrolling at programs in the previous year. Recalling our theoretical test of long-run stability if and only if the distribution of student types converge, we study the evolution of program PYSs over time in the Australian market. We show that volatility in program PYS decreases with time, and almost entirely stabilizes by the twelfth year that we observe a program in the data. However, there is significant entry and exit into the market, which does not allow all program PYSs to reach steady state. We show theoretically and empirically that programs with high PYSs are unaffected by entry and exit of programs with lower PYSs, and we note empirically that the entry and exit of programs typically happens amongst those with lower PYSs. We show that this implies that there is long-run stability at the "top" of the market, but not at the bottom. We show that this instability is associated with higher attrition rates, and also, that the instability particularly impacts low socio-economic status students.

Finally, we study the New York City high school admissions. We model the main takeaway of Abdulkadiroğlu et al. (2020)–that students prefer schools with higher achieving peers, and that the peer distribution is a sufficient statistic for overall student preferences. In this market,

[7] Nei and Pakzad-Hurson (2019) also discuss how learning new information affecting preferences can impact the stability of a higher-education market.

we show that a stable matching is generated in every period–with or without entry and exit of programs.

The overall approach of our study compares and contrasts to the well-known attempts to handle the presence of couples, who have preferences over pairs of positions (i.e. they often wish to be geographically colocated), in labor matching markets. Alvin E. Roth assisted in redesigning the National Residency Matching Program (NRMP), matching recent medical school graduates to hospitals in the late 1990s. As in our situation, the initial matching mechanism did not allow participants to fully explain their preferences. As Roth (2002, p. 1155) writes:

> To state the matter starkly, none of the conclusions of Theorems 1–4 apply to the medical match, not even that a stable matching always exists... What makes the NRMP different from a simple market is that it has complications. ...[C]ouples... need a pair of positions, and individual applicants also need two positions, because they match to positions for second year graduates, and then need to find a prerequisite first year position.

Roth and Peranson (1999) further state that "many of the existing theorems rest on assumptions not met in the complex medical market, and many of the medical market's complexities are known to open the door to the possibility of serious design problems." Roth and Peranson (1999) go on to deal with this misspecification by creating a mechanism that always finds a stable matching with couples when one exists. Kojima, Pathak, and Roth (2013) show that a stable matching exists when the share of couples is small in a large market, an assumption which appears justified given that only 5% of participants in the NRMP were members of couples.

Although both couples and the version of peer preferences that we study are due to preference complementarities, the former depends crucially on the identity of an individual in the market (one's partner) whereas the latter depends on the overall distribution of student ability. The approach of Roth and Peranson (1999) in redesigning the matching mechanism is also different from ours, where we study a smaller modification to the classical mechanism, namely an increase in transparency regarding the previous cohort's matching. There is also a difference in scope: while a small fraction of participants in the NRMP are members of couples, we identify many students in our setting as having peer preferences. Indeed, we show that the average student has a 20% chance of enrolling at a different program if she did not have peer preferences.

The paper is structured as follows: Section II introduces our model and main theoretical results; Section III presents the Australian Tertiary Education System and provides evidence of peer preferences; Section IV analyzes whether two markets, Australian college admissions and New

York City high school admissions achieve stability in the presence of peer preferences; Section V concludes. Omitted proofs and additional results are relegated to the A.1.

II Model

II.A. Setup

Our setup is drawn from those of Azevedo and Leshno (2016) and Leshno (2021). We initially discuss a static environment, and later move to a dynamic one with a new cohort of students to be matched in each period.

A continuum of students is to be matched to a finite set of programs $C = \{c_1, c_2, ..., c_N\} \cup \{c_0\}$. c_0 represents the outside option for each student. Each program $c \in C$ has capacity $q^c > 0$ measure of seats, with $q^{c_0} = \infty$. Let $q = \{q^c\}_{c \in C}$. Θ represents the set of student types, with typical element θ. η is a non-atomic measure over Θ in the Borel $\sigma-$algebra of the product topology of Θ, and H is the set of all such measures. We normalize $\eta(\Theta) = 1$ for all $\eta \in H$.

We begin by defining an assignment of students to programs. An *assignment* α is a measurable function $\alpha : C \cup \Theta \to 2^{\Theta} \cup 2^{C}$ such that:

1. for all $\theta \in \Theta, \alpha(\theta) \subset C$,

2. for all $c \in C$, $\alpha(c) \subset \Theta$ is measurable, and

3. $\theta \in \alpha(c)$ if and only if $c \in \alpha(\theta)$.

Condition 1. states that a student can be assigned to any subset of programs, condition 2. states that a program can be assigned to any subset of students, and condition 3. states that a student is assigned to a program if and only if the program is also assigned to that student. Note that this definition does not take into account feasibility. Specifically, it does not rule out situations in which capacity constraints are violated–a student can be assigned to multiple programs, and a program can be assigned to a larger measure of students than its capacity. Therefore, not all assignments are feasible, but this construction is a useful building block. Let $\mathcal{A}$ be the set of all assignments.

We further restrict assignments to take into account feasibility. A *matching* μ is a measurable function $\mu : C \cup \Theta \to 2^{\Theta} \cup C$ such that:

1. for all $\theta \in \Theta, \mu(\theta) \in C$,

2. for all $c \in C$, $\mu(c) \subset \Theta$ is measurable and $\eta(\mu(c)) \leq q^c$, and

3. $\theta \in \mu(c)$ if and only if $c = \mu(\theta)$.

Compared to an assignment, Condition 1 adds that a student can only be matched to one program, and condition 2 adds that the measure of students matched to a program cannot exceed the capacity of that program. We will often refer to a student θ for whom $\mu(\theta) = c_0$ as being "unmatched." Let $\mathcal{M}$ be the set of all matchings.

Each student type $\theta \in \Theta$ is given by $\theta = (u^\theta, r^\theta)$. $u^\theta(c|\alpha)$ represents the cardinal utility θ derives from being assigned to only program c given that other students are assigned according to assignment $\alpha \in \mathcal{A}$. That is, $u^\theta(c|\alpha) = u^\theta(c|\alpha(\theta) = c \text{ and } \{\alpha(\theta')\}_{\theta' \in \Theta \setminus \{\theta\}})$. As any matching restricts that a student θ cannot be matched to multiple programs, u^θ is sufficient to fully describe preferences for our purposes. We normalize $u^\theta(c_0|\alpha) = 0$ for all $\theta \in \Theta$ and $\alpha \in \mathcal{A}$, that is, each student receives a constant utility from being unassigned regardless of the assignments of other students. $r^{\theta,c} \in [0,1]$ is student θ's score at program c. We write r^θ to represent the vector of scores for student θ at each program. As scores will only convey ordinal information in our analysis, Without loss of generality, we assume that for each $\eta \in H$, $\eta\{\theta | r^{\theta,c} < x\} = x$ for all $x \in [0,1]$ and all $c \in C$, i.e. that the marginal distribution of every program's rankings is uniform.

We denote an economy by $E = [\eta, q]$, a distribution of student types and a vector of program capacities.

It will often be useful to denote the ordinal preferences of $\theta \in \Theta$ induced by u^θ. Let $\mathcal{P}$ be the set of all possible linear orders over programs $c \in C$. Let $\succeq^{\theta|\alpha} \in \mathcal{P}$ represent θ's induced preferences over programs at assignment α, that is $c_i \succeq^{\theta|\alpha} c_j$ ($c_i \succ^{\theta|\alpha} c_j$) if and only if $u^\theta(c_i|\alpha) \geq u^\theta(c_j|\alpha)$ ($u^\theta(c_i|\alpha) > u^\theta(c_j|\alpha)$).

To capture that peer preferences depend on the "ability" of students at a program, we consider the distribution of scores at each program given an assignment. For each $x \in [0,1]^{N+1}$, $c \in C$, and $\alpha \in \mathcal{A}$, let $\lambda^{c,x}(\alpha) := \eta(\{\theta | r^{\theta,c} \leq x \text{ and } \theta \in \alpha(c)\})$. Let $\lambda^c(\alpha)$ be the resulting non-decreasing function from $[0,1]^{N+1}$ to $[0,1]$ and let Λ be the set of all such functions.[8] Let $\lambda(\alpha) := (\lambda^{c_1}(\alpha), ..., \lambda^{c_N}(\alpha), \lambda^{c_0}(\alpha))$. In words, $\lambda(\alpha)$ represents the vector of ability distributions at each program for assignment α.

We now make a number of assumptions both to remove nuisance cases and to better reflect our desired environment.

A1 Scores and preferences are strict: for any $\theta \in \Theta$ and $c \in C$, $\eta(\{\theta' \in \Theta | r^{\theta'} = r^\theta\}) = 0$. For any $\alpha \in \mathcal{A}$, $\eta(\{\theta | \succeq^{\theta|\alpha} \text{ is strict}\}) = 1$.

A2 Full support for all α: Let $R \subset [0,1]^{N+1}$ be the support of scores induced by η, that is, R is the set of score vectors r such that for all $\epsilon > 0$, $\eta(\{\theta \in \Theta | \epsilon > ||r - r^\theta||_\infty\}) > 0$. Let $B_r(\epsilon)$ be the set of points within ϵ distance of $r \in R$, $B_r(\epsilon) := \{r' \in [0,1]^{N+1} | \epsilon > ||r - r'||_\infty\}$. Then for

[8] We endow this space with the pointwise convergence topology.

any $\alpha \in \mathcal{A}$, any $c, c' \in C \setminus \{c_0\}$, and any $r \in R$, $\eta(\{\theta \in \Theta | r^\theta \in R \cap B_r(\epsilon) \text{ and } c \succ^{\theta|\alpha} c'\}) > 0$.

A3 Student preferences depend only on $\lambda(\alpha)$, that is, for any $\alpha \in \mathcal{A}$ and any $\theta \in \Theta$, $\succeq^{\theta|\alpha} = \succeq^{\theta|\lambda(\alpha)}$.

We restrict our focus to economies E satisfying regularity conditions **A1**-**A3**. Additionally, we will assume the following regularity condition for certain results:

A4 Peer preferences are continuous, that is, for any $\epsilon > 0$ there exists some $\delta > 0$ such that if for any two assignments $\alpha, \alpha' \in \mathcal{A}$ we have that $\sup_{c,x} |\lambda^{c,x}(\alpha) - \lambda^{c,x}(\alpha')| := ||\lambda(\alpha) - \lambda(\alpha')||_\infty < \delta$, then $\eta(\{\theta \in \Theta | \succeq^{\theta|\alpha} \neq \succeq^{\theta|\alpha'}\}) < \epsilon$.

A2 and **A4** are richness assumptions: **A2** assumes that for any student, there exist other students with similar scores who have arbitrarily different preferences; **A4** assumes that the ordinal rankings of the vast majority of students do not change for small changes in the composition of peers.

Before introducing our desired solution concept of a stable matching, we discuss a multiplicity caused by the continuum assumption. To reduce a multitude of essentially-identical matchings which differ only for a zero-measure set of students, we only consider matchings $\mu \in \mathcal{M}$ that are *right continuous*: for any c and θ, if $c \succ^{\theta|\mu} \mu(\theta)$ then there exists $\epsilon > 0$ such that $\mu(\theta') \neq c$ for all θ' with $r^{\theta',c} \in [r^{\theta,c}, r^{\theta,c} + \epsilon)$.

A student-program pair (θ, c) *blocks* matching μ if $c \succ^{\theta|\mu} \mu(\theta)$ and either (i) $\eta(\mu(c)) < q^c$, or (ii) there exists $\theta' \in \mu(c)$ such that $r^{\theta,c} > r^{\theta',c}$. In words, θ and c block matching μ if θ prefers c to her current program (given peer preferences at μ) and either c does not fill all of its seats, or it admits a student it ranks lower than θ. A matching is *stable* if there do not exist any student-program blocking pairs.[9]

We build the tools to characterize stable matchings based on Azevedo and Leshno (2016), by first characterizing a class of assignments defined by admission cutoffs. Cutoffs are formally defined as arbitrary vectors $p \in \mathbb{R}_+^{N+1}$, subject to $p^{c_0} = 0$. One can construct an assignment for a given vector of cutoffs p in the following way. First, fix an arbitrary assignment α', and corresponding ability distribution $\lambda = \lambda(\alpha')$. Second, let each student type θ choose her favorite program among those where her program-specific score is weakly above the cutoff.[10] This program is called the **demand** of student θ, and is denoted by

[9] Note that the definition of stability implicitly assumes the feasibility conditions required by the definition of a matching, that is, that a student is matched to only one program and that a program does not have a larger measure of assigned students than its capacity.

[10] By assumption **A1**, a measure zero set of students may not have a single favorite program, and may be indifferent between several programs. In this case let the student break ties arbitrarily, and in what follows let $D^\theta(p, \lambda) \in \arg\max_{\succeq^{\theta|\lambda}} \{c \in C | r^{\theta,c} \geq p^c\}$. As this is relevant for at most a measure zero set of students, we proceed as if each student has a unique top choice for notational simplicity.

$$D^{\theta}(p,\lambda) = \underset{\succeq^{\theta|\lambda}}{\arg\max}\{c \in C | r^{\theta,c} \geq p^c\}$$

The fact that $p^{c_0} = 0$ means that any student can be unmatched.

We similarly define the demand for program c is given by :

$$D^c(p,\lambda) = \eta(\{\theta | D^{\theta}(p,\lambda) = c\})$$

The assignment $\alpha = A(p,\lambda)$ is defined by setting $\alpha(\theta) = D^{\theta}(p,\lambda)$ for every $\theta \in \Theta$. By construction, each student θ is assigned to exactly one program in assignment $\alpha = A(p,\lambda)$, but a program may be assigned to a larger measure of students than its capacity. As we are interested in characterizing (stable) matchings through cutoff vectors and score distributions (p,λ), we present the following two conditions on (p,λ). As we show, the first condition alone ensures that $A(p,\lambda)$ is a matching, and both conditions together ensure that $A(p,\lambda)$ is a stable matching.

Definition 1. *A pair (p,λ) of cutoffs and score distributions is* market clearing *if for all programs $c \in C$ we have*

$$D^c(p,\lambda) \leq q^c$$

and $p^c = 0$ when the inequality is strict.

Lemma 1. *If a pair (p,λ) is market clearing, then $A(p,\lambda)$ is a matching.*

The proof of this result is immediate, as for each $c \in C$, $\eta(\alpha(c)) \leq q^c$ and for each $\theta \in \Theta$, $\alpha(\theta) \in C$. If (p,λ) is market clearing, we often refer to matching $\mu = A(p,\lambda)$ as being *market clearing*, and we denote by M the set of all market clearing matchings, that is $M = \{\mu | \mu = A(p,\lambda)$ for some market clearing $(p,\lambda)\}$. By construction, $M \subset \mathcal{M}$.

Definition 2. *A pair (p,λ) represents* rational expectations *if it induces an assignment $\alpha = A(p,\lambda)$ such that $\lambda = \lambda(\alpha)$.*

The following lemma, a direct corollary of the supply and demand lemma of Azevedo and Leshno (2016) and Leshno (2021) holds:

Lemma 2. *If a pair (p,λ) is market clearing and rational expectations, then $\mu = A(p,\lambda)$ is a stable matching. Define $\hat{p}^c := inf\{r^{\theta,c} | \theta \in \mu(c)\}$ and let $\hat{p} = (\hat{p}^1, ..., \hat{p}^N, 0)$. If μ is a stable matching, then $(\hat{p},\lambda)$ are market clearing and represent rational expectations for $\lambda = \lambda(A(\hat{p},\lambda(\mu))$.*

The following result tells us that a stable matching exists in a large class of economies. Our proof extends the technique of Leshno (2021).

Theorem 1. *There exists a stable matching in any economy E satisfying **A4**.*

In contrast to the standard model without peer preferences, the set of stable matchings need not be unique.

Remark 1. *The set of stable matchings is not in general a singleton.*

We show this via the following example. In it, there are sufficiently many students who have strong peer preferences and desire classmates with higher scores, so that the "best" program is endogenously determined by the coordination of top-scoring students.

Example 1. *There are two programs, c_1 and c_2, where $r^{\theta,c} = r^{\theta}$ for $c \in \{c_1, c_2, c_0\}$. Student scores r^{θ} are distributed uniformly over $[0,1]$. Both programs have identical capacities $q^{c_1} = q^{c_2} < \frac{1}{2}$. For $i \in \{1,2\}$ let*

$$s^{c}(\lambda) = \frac{1}{\lambda^{c,(1,1,1)}} \int_0^1 y d\lambda^{c,(y,y,y)}$$

that is, $s^{c_i}(\lambda)$ is the mean score of students matched to c_i in μ.

For any $\lambda = (\lambda^{c_1}, \lambda^{c_2})$, all students prefer to be matched to any of the programs to being unmatched. Most students prefer to have peers with higher scores, but there is a 2ϵ measure of students who have "weak peer preferences," where $\epsilon \in (0, \frac{1}{2}]$: an ϵ measure of students who prefer c_1 to c_2 for any λ and an ϵ measure of students who prefer c_2 to c_1 for any λ, where these students are "uniformly distributed" in the skill distribution, i.e. the measure of students who have weak peer preferences and prefer program c_i with scores in interval (a,b) is $b-a$. The remaining students have strong peer preferences, and strictly prefer c_i to c_j if $s^{c_i}(\lambda) - s^{c_j}(\lambda) > \frac{q}{2}$ and $\lambda^{c_i,(1,1,1)} > \epsilon$ for each $i \in \{1,2\}$. This example is consistent with our regularity conditions.[11]

Let

$$p^{c_i} = 1 - \frac{q}{1-\epsilon} \qquad , \qquad p^{c_j} = 1 - 2q$$

and

$$\lambda^{c_i,(y,y,y)} = \begin{cases} 0 & \text{if } y < p^{c_i} \\ (1-\epsilon)(y - p^{c_i}) & \text{if } y \geq p^{c_i} \end{cases} \quad , \quad \lambda^{c_j,(y,y,y)} = \begin{cases} 0 & \text{if } y < p^{c_j} \\ y - p^{c_j} & \text{if } y \in [p^{c_j}, p^{c_i}] \\ \frac{q - 2q\epsilon}{1-\epsilon} + \epsilon(y - p^{c_i}) & \text{if } y > p^{c_i} \end{cases}$$

[11] Note that we have only specified student ordinal preferences in the case that the mean scores of students at the two programs are sufficiently different, meaning there are many utility functions that satisfy our regularity assumptions and comport with this example. Although we have defined peer preferences in terms of s and not λ, $s^{c_i}(\cdot)$, $i \in \{1,2\}$ is continuous in λ and therefore the continuous mapping theorem implies that assumption **A4** is satisfied.

Note that given our assumption that $\epsilon \leq \frac{1}{2}$, $p^{c_i} \leq p^{c_j}$. Let $p = (p^{c_i}, p^{c_j})$, $p' = (p^{c_j}, p^{c_i})$, $\lambda = (\lambda^{c_i}, \lambda^{c_j})$, and $\lambda' = (\lambda^{c_j}, \lambda^{c_i})$. We claim that $\mu = A(p, \lambda)$ and $\mu' = A(p', \lambda')$ are both stable matchings for sufficiently small ϵ. To see this, note that (p, λ) is market clearing because all students with scores weakly above p^{c_i} (except for those who have weak peer preferences and intrinsically prefer c_2) prefer to attend c_1 and all remaining students with scores weakly above p^{c_j} prefer to attend c_2. (p, λ) is continuous in ϵ, and as $\epsilon \to 0$, $s^{c_1} - s^{c_2} \to q > \frac{q}{2}$. Given our assumption on peer preferences, this implies (p, λ) represents rational expectations for sufficiently small ϵ, that is, all students with strong peer preferences will prefer c_1. Therefore, there is some $\epsilon^* > 0$ such that for all $\epsilon < \epsilon^*$, μ is stable. Leveraging symmetry, an analogous argument implies that μ' is also stable for all $\epsilon < \epsilon^*$.

II.B. Canonical Mechanisms and Stability

Theorem 1 tells us that a stable matching exists in a broad class of economies. Can a market maker ensure one using one of the "canonical" matching mechanisms typically studied in the literature? The following result suggests that the answer may be "no" in many settings.

Following Lemma 2, a stable matching must be market clearing and satisfy rational expectations. It may be natural to expect a matching to be market clearing in a canonical mechanism: no program can enroll more students than capacity and students apply to their preferred programs. However, rational expectations are more dubious. In order to generate a stable matching, students either need to correctly anticipate the ability distribution at each program (which may require common knowledge of student types, computational sophistication, and coordination if the set of stable matchings is not a singleton), or a matchmaker must be able to solicit student preferences over programs contingent on peer-ability.

For a given economy E, define a *one-shot matching mechanism* φ as a simultaneous-move, deterministic game in which each student θ submits a strict order $\tilde{\succ}^{\theta}$ over programs. φ maps submitted preferences $\tilde{\succ} = \{\tilde{\succ}^{\theta}\}_{\theta \in \Theta}$ and scores into a matching, that is $\varphi : (\mathcal{P} \times [0,1]^{N+1})^{\Theta} \to \mathcal{M}$. In an abuse of notation, we represent the resulting matching from report $\tilde{\succ}$ as $\varphi(\tilde{\succ})$, the matching for student type θ as $\varphi^{\theta}(\tilde{\succ})$, and the matching for program c as $\varphi^{c}(\tilde{\succ})$.

We now state two properties on mechanisms that are frequently satisfied by canonical mechanisms. Note that these are both conditions involving submitted rankings $\tilde{\succ}$. A mechanism φ *respects rankings* if for any $\tilde{\succ}$, $r^{\theta,c} \geq r^{\theta',c}$ for all c implies that $\varphi^{\theta}(\tilde{\succ}) \tilde{\succeq}^{\theta} \varphi^{\theta'}(\tilde{\succ})$. A mechanism respects rankings if it assigns a student type θ with higher scores at all programs than another student type θ' to a program that is ranked no lower (according to $\tilde{\succ}^{\theta}$) than the matching of student θ'. A mechanism φ is *stable* if for any $\tilde{\succ}$, $\varphi(\tilde{\succ})$ is stable *with respect to* $\tilde{\succ}$. A mechanism is stable if it creates a stable matching with respect to the submitted preferences. Note that any stable mechanism φ must respect rankings.[12]

The following result says that we can expect a clearinghouse to generate a stable matching if students have full knowledge of the distribution of student types.[13] In this case, the set of stable matchings is Nash-implemented by any stable mechanism φ as students are able to "roll in" peer considerations into their ordinal rankings over programs. That is, for any stable matching μ_*, there is an equilibrium in which each student type θ reports $\succsim^\theta = \succeq^{\theta|\mu_*}$.[14]

On the other hand, if students do not have full knowledge of the distribution of types, then we should not necessarily expect a clearinghouse to generate a stable matching. We represent the beliefs of $\theta \in \Theta$ over measures as $\sigma^\theta \in \Delta H$. Let $\succsim$ be a strategy profile, and in an abuse of notation, let $\succeq^{\theta|\sigma,\succsim}$ represent θ's expected ordinal rankings over programs given σ and $\succsim$. We say that student type θ *lacks rationality for the top choice at* $\succsim$ if the $\succeq^{\theta|\sigma,\succsim}$-maximal program is not the same as the $\succeq^{\theta|\varphi(\succsim)}$-maximal program.[15] For any $r \in [0,1)^{N+1}$ let $L_{\succsim,r} := \{\theta | r^\theta \geq r$ and θ lacks rationality for the top choice at $\succsim\}$.

Proposition 1. *Consider a one-shot matching mechanism φ.*

1. *Let φ be stable. Then the set of all stable matchings of economy E is identical to the set of all Nash equilibrium outcomes of φ.*

2. *Let φ respect rankings and let μ_* be a stable matching. If for any $r \in [0,1)^{N+1}$ and any $\succsim$ it is the case that $\eta(L_{\succsim,r}) > 0$ then there is no (Bayes) Nash equilibrium of φ that generates μ_*.*

The presence of some students with incorrect beliefs is not necessarily enough to lead to an unstable matching; in order for the presence of such students to affect the resulting matching, a number of conditions must be met. First, these students must have sufficiently strong peer preferences that their incorrect beliefs change their ordinal preferences over programs, for otherwise, their stated preferences would not change. Second, at least some of these students must have sufficiently high scores at programs, as the reported preferences of students with scores too low to match to any program will not affect the final outcome. Third, the incorrect beliefs must affect

[12] Proof: Suppose not. Then for some $\succsim$ there exist θ, θ' with $r^{\theta,c} \geq r^{\theta',c}$ for all c and $c^* = \varphi^{\theta'}(\succsim)\succsim^\theta\varphi^\theta(\succsim)$. But then $r^{\theta,c^*} \geq r^{\theta',c^*}$, implying that (θ, c^*) form a blocking pair. Contradiction with φ being stable.

[13] Full knowledge of the distribution of types is not a necessary condition for the clearinghouse to generate a stable matching. A stable matching can be generated in equilibrium if student beliefs are sufficiently close to the truth, which uses a similar logic to the convergence results we develop in Section II.C.

[14] Moreover, as our constructive proof shows, for any stable matching μ_* and all θ, there exists an equilibrium $\succsim^\theta$ which lists only one program as acceptable such that $\varphi(\succsim) = \mu_*$; even if there is a cap on the number of programs that students can list, which is common in many school-choice markets around the world, stable matchings can be generated under full rationality.

[15] Our result will not depend on the zero measure set of students who potentially have two top-ranked programs, and therefore, we can break ties arbitrarily.

the preferences at the "top" of some students' rankings, because, for example, changes in the ranking order of programs that are deemed unacceptable do not affect the final matching. Informally speaking, these conditions are likely satisfied if students have a sufficiently rich set of beliefs.

II.C. Tâtonnement with Intermediate Matching and belief updating

Given Proposition 1, an important question is how students form beliefs over the matching to be generated via a centralized mechanism. We model belief formation in a tâtonnement-like process, in which beliefs over the resulting matching are updated given the matching created by the previous cohort of students matched to programs.

Consider a discrete-time, infinite horizon model, where at every time $t = 1, 2, 3, ...$, the same programs are matched to a new cohort of students. For any $t, t' \geq 1$, economies E_t and $E_{t'}$ are identical, that is, the distribution of student types and program capacities are constant over time. We therefore omit all time indices when denoting student types $\theta \in \Theta$ and capacity vector q.

We describe the following dynamic matching process, which we call *Tâtonnement with Intermediate Matching (TIM)*. At each time period $t \geq 1$, a matching μ_t is constructed in the following way:

The market is initialized with an arbitrary $\mu_0 \in \mathcal{A}$ (our results would be qualitatively unchanged if we instead allowed students to have (potentially heterogeneous) beliefs over the initial assignment μ_0, but the exposition would become more cumbersome with this additional generality).[16] Incoming students at time t observe μ_{t-1}. A centralized matchmaker solicits rank-order lists of students, and then uses a stable matching mechanism to construct matching μ_t. We assume (and later show empirical evidence that) students use information from the previous period in a Cournot-updating fashion, that is, period t students assume that the matching μ_t will equal μ_{t-1} and they submit a rank-order list that best responds to μ_{t-1}. In an abuse of terminology, we often refer to μ_t, $t \geq 0$ as a matching, despite the fact that μ_0 is only required to be an assignment.

An instructive observation moving forward is that, assuming each student θ reports $\succeq^{\theta|\mu_{t-1}}$, μ_t is the unique stable matching in an economy where preferences are defined by μ_{t-1}. Formally, define measure $\zeta^{\eta,\alpha}$ as follows: for any open set $R \subset [0,1]^{N+1}$, any assignment $\alpha \in \mathcal{A}$ and any $\succeq \in \mathcal{P}$, $\zeta^{\eta,\alpha}(\{\theta | r^\theta \in R \text{ and } \succeq^{\theta|\alpha'} = \succeq\}) = \eta(\{\theta | r^\theta \in R \text{ and } \succeq^{\theta|\alpha} = \succeq\})$ for all $\alpha' \in \mathcal{A}$. In words, $\zeta^{\eta,\alpha}$ fixes the ordinal preferences of each student as they are in the original market for assignment α.

Remark 2. *In an economy $E = [\eta, q]$ let μ_t be the matching at time period $t \geq 1$ in the TIM process. Then μ_t is the unique stable matching in the economy $E' = [\zeta^{\eta,\mu_{t-1}}, q]$.*

This result follows from assumption **A2** and Theorem 1 of Azevedo and Leshno (2016). It

[16] The lone exception is that in Proposition 3, a stable matching would be generated in all periods, with the possible exception of period 1.

implies that μ_t is the outcome of student proposing deferred acceptance in the TIM procedure which is a strategy-proof mechanism in the static setting.[17] Therefore, we adopt this assumption that each student θ will submit her "true" preferences $\succeq^{\theta|\mu_{t-1}}$ in *any* stable matching mechanism.

Each μ_t is associated with a vector (p_t, λ_t) where p_t is the (unique) market-clearing cutoff vector given μ_{t-1}, and $\lambda_t = \lambda(A(p_t, \lambda_{t-1}))$.[18] Note that the entire sequence of TIM matchings $\{\mu_t\}_{t\geq 1}$ is uniquely determined by μ_0.

We can formulate the TIM process through two operators. The first is $P : \Lambda^{N+1} \to [0,1]^{N+1}$, which takes an ability distribution vector λ' and maps it into the (unique) cutoff vector that clears the market given λ', that is, $P\lambda' = p$ such that (p, λ') is market clearing.

The second is $S : [0,1]^{N+1} \times \Lambda^{N+1} \to \Lambda^{N+1}$ which outputs an ability distribution vector λ for each program in the present period's market assignment, that is, $S(p, \lambda') = \lambda(A(p, \lambda'))$.

Given an exogenous μ_0 resulting from μ_0, the dynamic process explained above tells us that $\mu_t = A(p_t, \lambda_{t-1})$, where $p_t = P\lambda_{t-1}$ and $\lambda_{t-1} = S(p_{t-1}, \lambda_{t-2})$. If $(p_{t-1}, \lambda_{t-1}) = (p_t, \lambda_t)$, the TIM process has reached steady state. This implies that $\mu_t = A(P\lambda_{t-1}, \lambda_{t-1}) = A(P\lambda_t, \lambda_t) = \mu_{t+1}$, i.e. the same matching is generated in all periods $t' \geq t$ in the TIM process. Note that if $\lambda_t = \lambda_{t-1}$, then $p_t = P\lambda_{t-1} = P\lambda_t = p_{t+1}$, and $\lambda_t = S(P\lambda_{t-1}, \lambda_{t-1}) = S(P\lambda_t, \lambda_t) = \lambda_{t+1}$. This implies that when the ability distribution vector reaches a steady state, the TIM process reaches a steady state in the following period.

The following result relates a steady state of the ability distribution vector (and therefore a steady state of the TIM process) to arrival at a stable matching. If and only if the ability distribution vector is in steady state does the TIM process generate a stable matching. Moreover, if and only if the ability distribution vector is in "approximate" steady state does the TIM process generate an "approximately" stable matching. Before stating the result, we formalize our notion of approximate stability.

Definition 3. *A matching μ is ϵ-*stable *if the measure of students involved in blocking pairs at μ is strictly smaller than ϵ, that is, $\eta(\{\theta|(\theta, c)$ block μ for some $c \in C\}) < \epsilon$.*

Theorem 2. *Let E be an economy, and let $\mu_1, \mu_2, ...$ be the sequence of matchings constructed in the TIM process given an initial μ_0.*

1. *λ_* is in steady-state if and only if the matching $\mu_* = A(P\lambda_*, \lambda_*)$ is stable.*

2. *For any $t \geq 1$ and any $\delta > 0$ there exists $\epsilon > 0$ such that if μ_t is $\epsilon-$stable, then $||\lambda_t - \lambda_{t-1}||_\infty < \delta$. Moreover, if E satisfies **A4** then for any $t \geq 1$ and any $\epsilon > 0$ there exists $\delta > 0$ such that if $||\lambda_t - \lambda_{t-1}||_\infty < \delta$, then μ_t is $\epsilon-$stable.*

[17] See Abdulkadiroğlu, Che, and Yasuda (2015) for further details on this mechanism in the continuum model.
[18] The uniqueness of p_t follows from Remark 2.

Consider an observer who does not necessarily know the preferences students have over peers, and only observes panel data on the ability distribution of entering classes at programs. This theorem provides a method for such an observer to analyze whether the market has (approximately) reached a stable matching. If and only if the ability distribution vector converges over time is the market "settling" into a stable matching.

Will the TIM procedure necessarily converge?

A natural question to ask is: does the TIM procedure always converge for any μ_0 in any economy E? If so, then the TIM process (approximately) delivers a stable matching in the long run. The first of the following two examples shows that the TIM procedure does not always converge. Moreover, the second example provides intuition for some of the features of a market that can lead to convergence.

Remark 3. *The TIM procedure does not necessarily converge, even when there is a unique stable matching.*

We show this result by counterexample, in an economy satisfying **A4**. First, we discuss lack of convergence, then uniqueness of the stable matching.

Example 2. *There is one program c (i.e. $N = 2$) with $q < 1$ measure of seats, and let $r^{\theta,c} = r^{\theta,c_0} = r^{\theta}$. Moreover, let $s(\lambda)$ represent the mean of scores of enrolled students at each program, that is,*

$$s(\lambda) = \frac{1}{\lambda^{c,(1,1)}} \int_0^1 y d\lambda^{c,(y,y)}$$

Each student θ receives zero utils from remaining unmatched, and receives utility $v^{\theta} - f(s(\lambda(\alpha)), r^{\theta})$ from matching with c at α, where

$$f(s(\lambda), r^{\theta}) = \begin{cases} 0 \text{ if } r^{\theta} \geq s(\lambda) \\ k \text{ if } r^{\theta} < s(\lambda) \end{cases}$$

The peer preference term $f(\cdot,\cdot)$ reflects that students want to be a "big fish" and suffer loss $k \in (0,1)$ if their score is not above average at the program. Therefore, a student θ is better off enrolling at c if and only if $v^{\theta} - f(s(\lambda), r^{\theta}) \geq 0$, where we break ties in favor of the student attending the program. Let each v^{θ} be distributed independently and uniformly over $[0,1]$.

Let $s_t = s(\lambda_t)$ for $t > 1$, and initialize the TIM procedure with μ_0 such that $s(\lambda(\mu_0)) \leq 1 - q$. Then $(p_1, s_1) = (1 - q, 1 - \frac{q}{2})$, as $\mu_1(c) = \{\theta | r^{\theta} \geq 1 - q\}$, that is, the top q mass of students enrolls at c because they expect (mistakenly for some) to face no peer loss from doing so.

What about (p_2, s_2)? Only the $1 - k$ fraction of students with $r^{\theta} < s_1$ for whom $v^{\theta} \geq k$ prefer enrolling in the program to remaining unmatched. All students with $r^{\theta} > s_1$ prefer to enroll in the program to being unmatched. Therefore, if the program fills all of its seats, then p_2 must solve

$$(1 - \frac{q}{2} - p_2)(1 - k) = \frac{q}{2}$$

or equivalently,

$$p_2 = 1 - \frac{q}{2} - \frac{q}{2(1-k)}$$

To simplify our analysis, we will deal only with the case in which the program fills all of its seats in μ_2 *(i.e.* $1 - \frac{q}{2} - \frac{q}{2(1-k)} \geq 0$*), which occurs if and only if* $k \leq 1 - \frac{q}{2-q}$*. Therefore, the average score of the "top half" of the students enrolled in the program is* $1 - \frac{q}{4}$ *while the average score of the "bottom half" of the students enrolled is* $\frac{1}{2}(1 - \frac{q}{2} + p_2)$*. This tells us that*

$$s_2 = \frac{1}{2}\left[1 - \frac{q}{4} + \frac{1}{2}(1 - \frac{q}{2} + p_2)\right]$$

When $k \geq \frac{4}{5}$*,* $s_2 \leq 1 - q$*.*[19] *But note then that* $(p_3, \lambda_3) = (p_1, \lambda_1)$*, as now all students with scores* $r^\theta > 1 - q$ *wish to enroll in the program. This creates a cycle wherein all even periods yield the same matching, while odd periods yield another (note that* $p_2 < p_1$*, as* $k > 0$*). Therefore, TIM does not converge.*

We now find cases in which the above economy has a unique stable matching. Assume, subject to later verification, that there exists a stable matching $\mu_* = A(p_*, \lambda_*)$ *in which c fills all of its seats. Let* $s_* = s(\lambda_*)$*. As all students* θ *with* $r^\theta \geq s_*$ *will attend c,* $1 - s_*$ *mass of seats are occupied by students who face no peer costs. In order for* p_* *to satisfy market clearing, it must be that* $(s_* - p_*)(1 - k) = q - (1 - s_*)$*, or equivalently that*

$$p_* = \frac{1 - q - ks_*}{1 - k}$$

As s is a function of λ*, a necessary condition for rational expectations of* (p_*, λ_*) *is that* $\frac{1+s_*}{2}(1 - s_*) + \frac{p_*+s_*}{2}(q - (1 - s_*)) = s_*$*. Solving this equation yields:*

$$s_* = \frac{k - kq - 2 \pm \sqrt{4 + k^2(q-1)^2 - 4k(q^2 - 3q + 1)}}{2k}$$

Noting that $k - kq - 2 < 0$*, only the "plus" solution is viable. In order for the "plus" solution to satisfy the necessary condition, it must be that* $(k - kq - 2)^2 \leq 4 + k^2(q-1)^2 - 4k(q^2 - 3q + 1)$*, which is shown, following a standard calculation, to hold with a strict inequality whenever* $q < 1$*.*

The above demonstrates that there is at most one stable matching in which c fills all of its seats. We argue that when q is sufficiently small any *stable matching must involve c filling all of its seats, by showing*

[19] Our simplifying assumption that the program fills all of its seats requires that $k \leq 1 - \frac{q}{2-q}$, which combined with the condition $k \geq \frac{4}{5}$, requires $q \leq \frac{1}{2}$.

that for sufficiently small q, it must be that $p_* > 0$. To see this, note that all students θ with $r^\theta > s_*$ will enroll in c. Therefore, $s_* > 1 - q$. For any fixed $k < 1$, $s_* \to 1$ as $q \to 0$. This implies that as $q \to 0$, $p_* = 0$ implies that $\eta(\mu_*(c)) \to 1 - k$, which violates the definition of matching as too large a measure of students is assigned to c.

By Theorem 1, there exists at least one stable matching, and our above arguments pin down the corresponding cutoffs p_* and average scores $s_* = s(\lambda(\mu_*))$ that must be identical in any two stable matchings for sufficiently small q. But if there exist two stable matchings, μ_* and $\mu_{\prime}$, note that by our assumption that student preferences depend on $s(\lambda)$, $\succeq^{\theta|\mu_*} = \succeq^{\theta|\mu_{\prime}}$ for all $\theta \in \Theta$. By Remark 2 it must be that $\mu_*(\theta) = \mu_{\prime}(\theta)$ for all $\theta \in \Theta$. Therefore, there is a unique stable matching for sufficiently small q.

To combine all of our conditions, we have shown that the TIM procedure does not converge given $s(\lambda_0) \leq 1 - q$ when $k \leq 1 - \frac{q}{2-q}$, and $k \geq \frac{4}{5}$. We have shown that there is a unique equilibrium when q is sufficiently small. Therefore, for any $s(\lambda_0) \in [0, 1)$ and any $k \in [\frac{4}{5}, 1)$ there exists $q' < 1$ such that for all $q > q'$ there exists a unique stable matching but the TIM procedure does not converge to it.

■

We now consider an example that is nearly identical to Example 2, and differs only in that $s(\lambda)$ represents the *median* of scores r^θ of enrolled students instead of the *mean* of the scores.

Example 3. *Consider Example 2 but where $s(\lambda)$ represents the median of scores r^θ of enrolled students at the program, that is, $s^c(\lambda) = \sup\{r | \frac{\lambda^{c,r}}{\lambda^{c,1}} \leq \frac{1}{2}\}$.*

Because of the assumption of uniformly distributed scores r^θ (and the intuitive similarities between the mean and median), the pair of cutoffs and median scores at $t = 1$ remains the same as in Example 2, given an upper bound on $s(\lambda_0)$: with $s(\lambda_0) \leq 1 - q$, $(p_1, s_1) = (1 - q, 1 - \frac{q}{2})$. Additionally, $p_2 = 1 - \frac{q}{2} - \frac{q}{2(1-k)}$. Note however that $s_2 = s_1 = 1 - \frac{q}{2}$; all of the students with scores $r^\theta \geq 1 - \frac{q}{2}$ "return" to the program, and while the set of students who attend the program with scores $r^\theta < 1 - \frac{q}{2}$ differs in periods 1 and 2, there are the same measure of them (filling exactly half of the seats), meaning that they do not affect the median. By our assumption that student preferences depend only on $s(\lambda)$, $\succeq^{\theta|\mu_1} = \succeq^{\theta|\mu_2}$ for all $\theta \in \Theta$. By Remark 2 it must be that $\mu_2(\theta) = \mu_3(\theta)$ for all $\theta \in \Theta$. Therefore, $\lambda(\mu_2) = \lambda(\mu_3)$ and by Theorem 2 TIM produces a stable matching for all $t \geq 2$.

■

The only difference between these two examples is that peer preferences depend on the mean of student scores in the former, and the median in the latter. The median is not affected by outliers: given that the top-ranked $\frac{q}{2}$ students enroll in the program for each $t \geq 1$, the median is guaranteed to stay the same in the TIM procedure. In contrast, the mean is sensitive to the entire distribution of enrolling students: even if the top-ranked $\frac{q}{2}$ students enroll, the mean can decrease if students with average rankings do not enroll and some with lower scores do. This can lead to a cycle and failure of convergence to stability.

II.D. Two Markets

In this section, we analyze the convergence, or lack there of, of the TIM procedure to a stable matching in two markets: the Australian college admissions market, and a public highschool market. We make assumptions to mirror key features of each market. We present empirical evidence to justify the assumptions for the Australian market in Section III and we defer to Epple and Romano (1998); Abdulkadiroğlu et al. (2020); Rothstein (2006); Beuermann and Jackson (2020) for the highschool markets.

An important consideration in these markets is that students likely have access to only a summary statistic of the distribution of peers in previous cohorts, not the entire distribution. We therefore briefly provide general theoretical results, mirroring those in the previous section, in markets where student preferences are based only on a summary statistic of the ability distribution. As the proofs follow straightforwardly from those of our original results, we omit them.

Definition 4. *For each $c \in C$ let a* summary statistic *of abilities at program c be a function $s^c : \Lambda \to [0,1]$. For $\lambda \in \Lambda^{N+1}$ let $s(\lambda) = \times_{c \in C} s^c(\lambda)$ be the vector of summary statistics.*

We provide the following regularity conditions, which roughly speaking subsume the roles of **A3** and **A4**.

A5 Student preferences depend only on $s(\lambda(\alpha))$, that is, for any assignment $\alpha \in \mathcal{A}$ and any θ, $\succeq^{\theta|\alpha} = \succeq^{\theta|s(\lambda(\alpha))}$.

A6 For any assignment α and $\epsilon > 0$ there exists some $\delta > 0$ such that if for an assignment α' we have that $||s(\lambda(\alpha)) - s(\lambda(\alpha'))||_\infty < \delta$, then $\eta(\{\theta| \succeq^{\theta|\alpha} \neq \succeq^{\theta|\alpha'}\}) < \epsilon$.

A7 For any matching $\mu \in M$ and $\epsilon > 0$ there exists some $\delta > 0$ such that if for a matching $\nu \in M$ we have that $||\lambda(\mu) - \lambda(\nu)||_\infty < \delta$, then $||s(\lambda(\mu)) - s(\lambda(\nu))||_\infty < \epsilon$.

We first provide an analogue to Theorem 1. We note that this result is similar to the existence result in Leshno (2021).

Corollary 1. *Let E be an economy satisfying* ***A1***, ***A2***, ***A5*** *−* ***A7***. *Then E has at least one stable matching.*

The following result mirrors Theorem 2. In an economy satisfying the required regularity conditions, an observer of the TIM process need only verify that the summary statistics of student abilities is in (approximate) steady state in order to determine that the market has (approximately) converged to stability.

Corollary 2. *Let E be an economy satisfying **A1**, **A2**, and **A5**, and let $\mu_1, \mu_2, ...$ be the sequence of matchings constructed in the TIM process for a given μ_0.*

1. *$s(\lambda_*)$ is in steady-state if and only if the matching $\mu_* = A(P\lambda_*, \lambda_*)$ is stable.*

2. *For any E satisfying **A6**, any $t \geq 1$ and any $\delta > 0$ there exists $\epsilon > 0$ such that if μ_t is $\epsilon-$stable, then $||s(\lambda_t) - s(\lambda_{t-1})||_\infty < \delta$. Moreover, if E satisfies **A7** then for any $t \geq 1$ and any $\epsilon > 0$ there exists $\delta > 0$ such that if $||s(\lambda_t) - s(\lambda_{t-1})||_\infty < \delta$, then μ_t is $\epsilon-$stable.*

II.D..1 The Australian Market

There are two important sets of stylized facts that our modeling of the Australian market attempts to match. First, students have "big-fish" preferences: each student has a one-dimensional ability that determines both university scores and peer preferences. Students suffer a utility loss if their score is below an *ordinal* summary statistic of the distribution of peers, but are indifferent toward their peers if their ability is above the summary statistic. Second, we relax our initial assumption that the market is identical in each period, and instead allow for changes due to the entry and exit of programs. In particular, more desirable programs are long-lived, but less desirable programs enter and exit the market over time.

In each period (i.e. for a given set of programs) we show that there is a unique stable matching in an Austrlian market. Moreover, we show that big-fish preferences are sufficient for convergence of the TIM process to this stable matching in the special case where there is no entry and exit of programs. With entry and exit, we show that only students with sufficiently high abilities are guaranteed to be matched to their stable partner in the long run. Therefore, there exists instability at the "bottom" of the market.

As before, let an economy be characterized by $E = [\eta, q]$ where $\eta \in H$ is the measure over student types Θ, and q is the capacity vector, where for each $c \in C = \{c_1, ..., c_N, c_0\}$, $q^c > 0$ and $q^{c_0} = \infty$. Let $E_1, E_2, ...$ be a sequence of economies, where for each $t \geq 1$ there is a set $C_t \subset C$ of active programs, where $|C_t| = N_t + 1$ and $c_0 \in C_t$. $E_t := [\eta, q_t]$ where $q_t = \times_{c \in C_t} q^c$ is the capacity vector for active programs $c \in C_t$. Let $\mathcal{A}_t$, $\mathcal{M}_t$, $A_t(p, \lambda)$, and M_t be the set of assignments in E_t, the set of matchings in E_t, the E_t market assignment for $(p, \lambda) \in [0,1]^{N_t+1} \times \Lambda^{N_t+1}$, and the set of all market clearing matchings in E_t, respectively.

We continue to assume that economy E satisfies **A1**, which implies that each economy E_t satisfies **A1**. We formalize the stylized restrictions on preferences with the following three points:

AA1 Common rankings: $r^\theta := r^{\theta,c} = r^{\theta,c'}$ for any $c \in C_t$, $c' \in C_t'$ with $t, t' \geq 1$, and $\theta \in \Theta$.

AA2 Big-fish preferences: For each $c \in C$, each $\theta \in \Theta$ has utility function $u^\theta(c|\alpha) = v^{\theta,c} -$

$f^{\theta,c}(r^{\theta}, s^{c}(\lambda(\alpha)))$, where $f^{\theta,c}(\cdot,\cdot) \geq 0$, is nondecreasing and continuous in its second argument, and $f^{\theta,c}(r^{\theta}, s^{c}(\lambda(\alpha))) = 0$ if $r^{\theta} \geq s^{c}(\lambda(\alpha))$.

AA3 k^{th} highest score: We say that $s^{c}(\cdot)$ represents the $(k^{c})^{th}$ *highest score* if there exists $k^{c} \in [0,1]$ such that for any market clearing matching $\mu \in M_t$, $s^{c}(\lambda(\alpha))$ equals the supremum value of r^{θ} for which $\eta(\{\theta' \in \mu(c) | r^{\theta'} > r^{\theta}\}) = k^{c}$ (if such a number exists, and 0 otherwise). For each $t \geq 1$ and each $c \in C_t$ there exists k^{c} where $s^{c}(\cdot)$ represents $(k^{c})^{th}$ highest score.

AA1 reflects the fact that a standardized score is used by programs for admission. AA2 states that students face an additive peer cost when assigned to a program where their score is below the summary statistic of the scores of their peers. AA3 represents that students have relative ranking concerns. An important part of AA3 is the restriction to the set of market clearing matchings M_t, but the restriction does not apply to other matchings. As a result, other functional forms of the summary statistic, including where $s^{c}(\cdot)$ represents the median score of students assigned to c can be accommodated for certain markets.[20]

The following reflects are stylized restrictions on entry and exit of programs. Let there be two disjoint "blocks" of programs $B_1 \subset C \setminus \{c_0\}$, and $B_2 \subset C \setminus \{c_0\}$ such that $B_1 \cup B_2 = C \setminus \{c_0\}$. To capture that more popular programs are longer lived, we additionally make the following three assumptions about student preferences over programs, and the entry and exit of programs.

AA4 Block one is always active: Every $c \in B_1$ is an element of C_t for every $t \geq 1$.

AA5 Block-correlated preferences: $u^{\theta,c} > u^{\theta,c'}$ for all $\theta \in \Theta$, all $c \in B_1$, and all $c' \in B_2$.

AA6 Full support: Let R be any open subset of $[0,1]$. Then for any $\alpha \in \mathcal{A}_t$ and any $c, c' \in B_1$, $\eta(\{\theta \in \Theta | r^{\theta} \in R \text{ and } c \succ^{\theta|\alpha} c'\}) > 0$.

AA4 captures that certain programs are long-lived. AA5 adds that long-lived programs are more desirable to students. This restriction is based on block-correlated preferences, discussed in Coles, Kushnir, and Niederle (2013). AA6 is a relaxation of A2, ensuring full support of preferences over the programs in the first block.

[20] The reason for the restriction of this definion to the set M_t is that the TIM procedure only produces matchings $\mu \in M_t$, therefore the sequence of matchings generated from two otherwise identical markets will be identical if their summary statistics vectors coincide on this restricted set of matchings. This means that a wider class of summary statistics fall into the category of k^{th} highest score than it might initially seem. Specifically, suppose $s^{c}(\cdot)$ represents the score of the $(100 \cdot m)^{th}$ percentile student assigned to c; for $m \leq 1$ let $s^{c}(\lambda(\alpha))$ equal the supremum value of r^{θ} for which $\eta(\{\theta' \in \alpha(c) | r^{\theta'} > r^{\theta}\}) = m \cdot \eta(\alpha(c))$. Then $s^{c}(\cdot)$ satisfies our definition of k^{th} highest statistic if for any two matchings $\mu, \nu \in M_t$, $\eta(\mu(c)) = \eta(\nu(c))$for all $c \in C_t \setminus \{c_0\}$. Since $\eta(\mu(c))$ does not vary in the set M_t, define $k^{c} := m \cdot \eta(\mu(c))$, and AA3 is satisfied. Therefore, summary statistics such as the median (see Example 3) can fit into the results of this section. Moreover, the condition that for any two matchings $\mu, \nu \in M_t$ we must have $\eta(\mu(c)) = \eta(\nu(c))$ is not "knife edge" (i.e. it holds for an open set of market fundamentals): suppose that for every $\theta \in \Theta$ and any $\alpha \in A_t$ it is the case that $c \succeq^{\theta|\alpha} c_0$ for all $c \in C$ and there is an undersupply of seats, $\sum_{c' \in C_t \setminus c_0} q^{c'} < 1$ for all $t \geq 1$. Then for all $\mu \in M_t$, and all $c \in C_t \setminus \{c_0\}$, $\eta(\mu(c)) = q^{c}$.

Definition 5. *We say that a sequence of economies $E, E_1, E_2, ...$ is* Australian *if it satisfies* ***A1***,***AA1***-***AA6***.

There exists a unique stable matching for each E_t, $t > 0$. Any student type θ with a sufficiently high scores in the stable matching for each market E_t, $t > 0$.

Proposition 2. *At any time $t > 0$ in an Australian market there exists a unique stable matching μ_t^*. Moreover,*

1. *For any $c \in B_1$ and any $c' \in C_t \cap B_2$, $s^c(\lambda(\mu_t^*)) \geq s^{c'}(\lambda(\mu_t^*))$,*

2. *For all $c \in B_1$, $s^c(\lambda(\mu_t^*)) = s^c(\lambda(\mu_{t'}^*))$ for all $t' \geq 1$, and*

3. *If there exists $c \in B_1$ such that $r^\theta \geq s^c(\lambda(\mu_t^*))$, then $\mu_t^*(\theta) = \mu_{t'}^*(\theta)$ for all $t' \geq 1$.*

We provide a pseudo-serial-dictatorship mechanism in the appendix which serves as a constructive proof of existence. To show uniqueness, suppose there are two stable matching summary statistic vectors in any E_t, $t \geq 1$. Take the program with the highest summary statistic among those with different statistics at the two matchings, s^{max}. All student types θ with scores r^θ greater or equal to s^{max} must have the same ordinal preferences over programs at the two matchings. But then it cannot be that the program fills k^c measure of seats with students with higher scores than s^{max} in one matching, but strictly fewer than k^c seats with students with higher scores than s^{max}. The logic from the remaining points of the proposition follow from our algorithm and proof of uniqueness.

The TIM procedure in this market with entry and exit is largely analogous to our base model. The market is initialized with an arbitrary assignment, and in each period, the unique market-clearing matching is constructed given the ability vector of the "incoming" assignment. The "incoming" assignment for any program active in both the current and previous periods is equal to that program's matching in the previous period, but due to entry and exit the "incoming" assignment for programs that were not active in the previous period is allowed to be arbitrary.

Formally, for each $t \geq 0$ there is an incoming assignment $\nu_t \in \mathcal{A}_{t+1}$. In each period $t \geq 1$ a matching $\mu_t \in M_t$ is formed as follows:

A time-dependent operator $P_t : \Lambda_t^{N_t+1} \to [0,1]^{N_t+1}$, maps an ability distribution vector λ' into the (unique) cutoff vector that clears market E_t given λ', that is, $P_t(\lambda') = p$ such that (p, λ') is market clearing in E_t. $\mu_t = A_t(P_t\lambda_{t-1}, \lambda_{t-1})$, where $\lambda_{t-1} = \lambda(\nu_{t-1})$.

$\nu_0 \in \mathcal{A}_1$ is an arbitrary assignment, and each subsequent assignment $\nu_t \in \mathcal{A}_{t+1}$ is constructed as follows: $\nu_t(c) = \mu_t(c)$ for all $c \in C_t \cap C_{t+1}$. For all $c \in C_{t+1} \setminus C_t$, $\nu_t(c)$ is arbitrary.

The following result matches our empirical findings. Regardless of entry and exit, the summary statistics of popular programs (those in block B_1) converge to their stable levels in the TIM procedure, and except in rare cases, this convergence occurs in finite time. Let $V := \{\theta \in \Theta | r^{\theta} \geq \min_{c \in B_1} s_*^c\}$ be the set of students with scores higher than the stable matching summary statistic of at least one program in block 1. We also find that all student types $\theta \in V$ eventually receive their stable matching partner.

Theorem 3. *In any Australian market the TIM procedure is such that there generically exists some time $T < \infty$ such that $s_t^c = s_*^c$ for all $c \in B_1$ and $\eta(\{\theta \in V | \mu_t(\theta) = \mu_*(\theta)\} = \eta(V)$ for all $t > T$.*

This result tells us that we should expect convergence of summary statistics for popular programs that do not see entry and exit (those in B_1), and moreover, sufficiently high-scoring students will receive their stable partner, regardless of entry and exit. However, it is not necessarily the case that the summary statistics of less popular programs that see entry and exit (those in B_2) converge–students with lower scores are not guaranteed to receive their stable partner in the long run. Therefore, instability only affects students with scores below the stable matching summary statistics of all programs in the top block.

The lowest-scoring students matched to top block programs may be affected by entry and exit, and there may not exist some $T > 0$ such that these students receive their stable matching program in all $t > T$. However, the "big fish" preferences present in Australian markets implies that even these students do not receive a negative utility shock to their preferences. We say that a student type $\theta \in \Theta$ is a member of a *negative utility blocking pair* if there exists $c \in C_t$ such that (θ, c) form a blocking pair, and $u^{\theta}(\mu_t(\theta)|\nu_{t-1}) > u^{\theta}(\mu_t(\theta)|\mu_t)$.

Remark 4. *In a generic sequence of economics $E, E_1, E_2, ...$ the measure of students involved in negative utility blocking pairs goes to zero in the TIM procedure, that is,*
$\eta(\{\theta \in \bigcup_{c \in B_1} \mu_t(c) | \theta$ is a member of negative utility blocking pair$\} \to 0$ Moreover, if $s_c^* = 0$ for at most one program $c \in B_1$, then there exists some time $T < \infty$ such that
$\eta(\{\theta \in \bigcup_{c \in B_1} \mu_t(c) | \theta$ is a member of negative utility blocking pair$\} = 0$ for all $t > T$ in the TIM procedure.

This result further solidifies the lack of stability at the "bottom" of the market: only students who are matched to programs in B_2 are potentially subject to a lower utility than anticipated from their program for sufficiently large t.

In the special case in which $B_1 = C \setminus c_0$, all programs are in the top block and there is therefore no entry and exit. Theorem 3 implies that the TIM procedure converges to the (unique) stable matching in finite time. Indeed, in this case, all of the general results we derive in Section II.D. hold, as shown by the following result.

Remark 5. *Let $C \setminus \{c_0\} = B_1$, and let $E = E_1 = ...$ be an Australian market. Moreover, for any $\alpha \in \mathcal{A}$ and any $c \in C$ let $k^c \in [0,1]$ be such that $s^c(\lambda(\alpha))$ equals the supremum value of r^{θ} for which*

*$\eta(\{\theta' \in \alpha(c) | r^{\theta'} > r^{\theta}\}) = k^c$ (if such a number exists, and 0 otherwise). Then E also satisfies **A2**, **A5**-**A7**.*

Given Remark 5, a relevant question is, "how much instability does entry and exit cause?" Consider the following thought experiment. Suppose that there is entry or exit of a new program(s) at period t, and that the set of programs remains constant until period $t + T$, where $T > 0$. If, starting at ν_t, it takes the TIM process fewer than T periods to converge, then $\mu_{t'}$ will be stable for periods $t' \in (t + T', t + T]$ for some $0 < T' < T$. If T' is much smaller than T, the market will generate a stable matching for a large fraction of the periods between the change in set of programs.

The following result upper bounds T' in this thought experiment, in markets where there is sufficient alignment in intrinsic student values over programs, $v^{\theta,c}$. For notational simplicity, we assume that $B_1 = C \setminus \{c_0\}$ and show that the TIM procedure converges from any starting condition μ_0 in no more than $N + 2$ periods. Therefore, if entry or exit happens far less often than once every $N + 2$ periods, the TIM procedure will "usually" generate a stable matching. This bound is tight; in the appendix, we show that there exist markets in which convergence does not occur in fewer than $N + 2$ periods.

Remark 6. *Let $B_1 = C \setminus \{c_0\}$. For any μ_0 and $\delta > 0$, there exists $\epsilon'^* > 0$ such that for any $0 < \epsilon < \epsilon'$, if the measure of students who have common intrinsic program preferences is strictly larger than $1 - \epsilon$, $\eta(\{\theta \in \Theta | v^{\theta,c_1} > v^{\theta,c_2} > ... > v^{\theta,c_N}\}) > 1 - \epsilon$, then μ_t is δ-stable for all $t > N + 1$.*

II.D..2 Pure Peer Preferences Markets

We now consider the case in which students prefer peers with higher ability, in contrast to our modeling of the Australian market. We adopt a common assumption introduced by Epple and Romano (1998): student preferences for a program are entirely based on the quality of peers at that program; Abdulkadiroğlu et al. (2020) provide supporting empirical evidence from New York City secondary schools that students (or their parents) have preferences for higher ability peers, and that after controlling for peer ability, there is no additional impact of school quality.

We model this by assuming students have preferences over programs that are increasing in the ability of peers at that program, without intrinsic payoffs for being assigned to any school in this market. We will refer to this as a *pure peer preferences economy*. As ability is measured using objective outcomes such as standardized test scores, we assume that student scores are the same for all programs, so that $r^{\theta,c} = r^{\theta}$ for all $c \in C$. Formally, for any assignment $\alpha \in \mathcal{A}$, and programs $c, c' \in C \setminus \{c_0\}$ and any student type $\theta \in \Theta$, if $\alpha(c) \neq \alpha(c')$ then either $c \succ^{\theta|\alpha} c'$ for almost all $\theta \in \Theta$ or $c' \succ^{\theta|\alpha} c$ for almost all $\theta \in \Theta$. Moreover, if for almost all $\theta \in \alpha(c)$ and almost all $\theta' \in \alpha(c')$ it is the case that $r^{\theta} > r^{\theta'}$, then $c \succ^{\theta|\alpha} c'$ for all $\theta \in \Theta$.[21] For simplicity of exposition, we assume that

[21] Note that any such economy does not satisfy assumption **A2**, to comport more closely with the conclusions of

for any assignment, all programs are acceptable for all students, that is, for any $\alpha \in \mathcal{A}$ and any program $c \in C \setminus \{c_0\}$, $c \succ^{\theta|\alpha} c_0$ for all $\theta \in \Theta$.

In a pure peer preferences economy, the TIM procedure generically generates a stable matching in all time periods $t \geq 1$.[22]

Proposition 3. *Let E satisfy pure peer preferences. Then for almost any $\mu_0 \in \mathcal{A}$, each matching μ_t, generated by the TIM process for $t \geq 1$ is stable.*

The proof is straightforward. Given any initial assignment μ_0 such that $\mu_0(c) \neq \mu_0(c')$ for any $c, c' \in C$, it will be the case that (almost) all students have the same ordinal preferences over programs at time $t = 1$. WLOG let $c_1 \succ^{\theta|\mu_0} c_2 \succ^{\theta|\mu_0} c_3 \succ^{\theta|\mu_0} ... \succ^{\theta|\mu_0} c_N$. Due to the common scores of programs, the top q^{c_1} scoring students will be matched to c_1 in μ_1, the next top q^{c_2} scoring students will be matched to c_2 in μ_1, and so on, until either all programs are full or all students are matched.

μ_1 is clearly stable; all students θ prefer to match with a lower-index program, but all such programs are filled to capacity with higher scoring students. Moreover, note that $\succeq^{\theta|\mu_0}=\succeq^{\theta|\mu_1}$ for almost all θ, as the most-desired program under μ_0, c_1, remains the most-desired program under μ_1, the second most-desired program under μ_0, c_2, remains the second most-desired program under μ_1, and so on. Therefore, $\mu_2 = \mu_1$ and is also stable. This logic holds for μ_t, $t \geq 1$.

Note that the stable matching generated in the TIM process is not unique. For a given μ_0, the ordinal preferences of students never change. Therefore, any permutation of programs leading to μ_0' will lead to a different stable matching.

The convergence of the TIM procedure to a stable matching in pure peer preferences economies varies from other economies in our paper. Unlike the economy studied in Example 1, the TIM procedure leads to a stable matching in a pure peer preference economy. Moreover, unlike in an Australian economy, this convergence to a stable matching happens immediately in a pure peer preferences economy.

An interesting implication of this immediate convergence is that even with exit and entry of programs, the TIM procedure creates a stable matching at every time $t \geq 1$.

II.E. A More Stable Mechanism

At least three problems exist with the TIM procedure. First, it need not converge, meaning that we are not guaranteed stability in the long run. Second, even if it does converge, initial cohorts will have unstable matchings if the convergence is not immediate. Third, as discussed

Abdulkadiroğlu et al. (2020). As in Example 1, slight adjustments could be made to this market to satisfy **A2**. Our conclusions in this section would not change.

[22] Under a similar assumption, Pycia (2012) finds that a stable matching always exists in a small, finite market.

in the previous section, there may be changes to the market from year to year, which potentially make convergence more difficult.

We present a mechanism that improves upon all three of these shortcomings of the TIM process. This mechanism does not run across years, and instead attempt to find or approximate a stable matching for each cohort of students. Unlike the TIM process, it suffers from neither instability before reaching steady state, nor instability caused by changes in the market over time. Moreover, as we show, it can yield an approximately stable matching even when the TIM process does not converge.

Each cohort is divided up into small subgroups where students are assigned to each subgroup “uniformly at random." Students in each subgroup report ordinal preferences sequentially after seeing the previous subgroup’s ability distribution. The mechanism operates more like a traditional tâtonnement process, in that no matching is created until the ability distribution vector (nearly) converges. Finally, both to create incentives for truthtelling and to reduce the measure of blocking pairs, we allow early participants in the mechanism to change their reported preferences to reflect the final ability distribution. Formalizing this idea involves specifying the student types in each of T submarkets, the programs (and measure of seats) in each submarket, and how peer preferences are defined relative to the original market. We use the subscript "t" below to be evocative of the time index in the TIM process presented above, but note that this mechanism is defined for a single market, and does not rely on dynamics across markets.

First, we specify the student types in each submarket. For any subset $\Theta_t \subset \Theta$, let η_t represent the induced measure over Θ_t. We partition Θ into sets Θ_1,...,Θ_T such that for each $t \in \{1, ..., T\}$, Θ_t is constructed “uniformly at random," i.e. for any $\theta \in \Theta_t$ and any open neighborhood $n(\theta) \subset \Theta$ of θ, it is the case that $\eta_t(n(\theta) \cap \Theta_t) = \eta(n(\theta)) \cdot \eta(\Theta_t)$. We assume that $\eta(\Theta_t) \to 0$ for all t as $T \to \infty$.

Second, we specify the programs. Each program $c \in C$ is active in each submarket, but has $q_t^c = q^c \cdot \eta(\Theta_t)$ seats available. We denote the entire vector of capacities in submarket t as q_t.

We use these measures and capacities to formally denote a submarket $t \in \{1, ..., T\}$ by $E_t = [\eta_t, q_t]$.

Third, we define the ability distribution. Let $\mathcal{A}_t$ be the set of all assignments in economy E_t. For each $x \in [0,1]^{N+1}$, $c \in C$, and $\alpha \in \mathcal{A}_t$ let the ability distribution in submarket t be denoted by $\lambda_t^{c,x}(\alpha) := \frac{\eta(\{\theta | r^\theta \leq x \text{ and } \theta \in \alpha(c)\})}{\eta_t(\Theta_t)}$. Let $\lambda_t^c(\alpha)$ be the resulting non-decreasing function from $[0,1]^{N+1}$ to $[0,1]$ and let Λ be the set of all such functions.[23] Let $\lambda_t(\alpha) := (\lambda_t^{c_1}(\alpha), ..., \lambda_t^{c_N}(\alpha), \lambda_t^{c_0}(\alpha))$.

The following mechanism creates a matching μ_t in each submarket, and iterates until near convergence of $\lambda_t(\alpha)$.

[23] We endow this space with the pointwise convergence topology.

Definition 6. *The* Tâtonnement with Final Matching (TFM) *mechanism is defined by the following steps:*

step 0: *Initialize the mechanism with $\delta > 0$, $T > 0$, and $\mu_0 \in \mathcal{A}$.*

step $\tau = K \cdot T + t$, $K \geq 0$, $t \in \{1, ..., T\}$: *Report to student types $\theta \in \Theta_t$ the distribution $\lambda(\mu_{\tau-1})$ and solicit their ordinal preferences over programs. Run (student proposing) deferred acceptance in sub-market E_t to create matching μ_τ.*

At the first step τ such that $\|\lambda(\mu_\tau) - \lambda(\mu_{\tau-1})\|_\infty < \delta$, terminate the process above. Show all student types $\theta \in \Theta$ distributions $\lambda(\mu_{\tau-1})$ and solicit their ordinal preferences over programs. Run (student proposing) deferred acceptance in the aggregate market E. The outcome of deferred acceptance in the aggregate market E is the final matching for all students.

For a given starting condition μ_0 and associated ability distribution $\lambda_0 = \lambda(\mu_0)$, the TFM mechanism depends on parameters δ and T. δ determines the final matching by defining the stopping criterion, and holding δ fixed, T determines how many times each subcohort is asked to report their preferences.

The following result states that the TFM mechanism converges if the TIM procedure does, and for sufficiently small δ creates a nearly-identical matching. Moreover, the TFM mechanism can create a nearly-stable matching even when the TIM procedure does not converge. We prove this by construction, which may be of independent interest–we show that the TIM procedure potentially suffers from a lack of local convergence; even if the TIM procedure creates a near stable matching in a particular time period t, it need not create a near-stable matching in subsequent periods (see Example 7 in the appendix). However, because the TFM mechanism terminates at any step such that the ability distribution vector is approximately steady, it creates a near-stable matching in such cases (see Theorem 2).

The TFM mechanism produces a nearly-stable matching with good incentive properties. We say that a student $\theta \in \Theta_t$ *misreports at step t* if she submits a preference profile $\tilde{\succsim}^\theta \neq \succsim^{\theta|\mu_{t-1}}$. We show that for any $\epsilon > 0$, there exists $\delta > 0$ defining the stopping rule such that no more than an ϵ measure of students can profitably misreport their preferences, assuming their peers do not themselves misreport. As the proof reveals, if we additionally assume that every student's cardinal preferences are continuous in λ,[24] then this point can be strengthened to show that there is an ϵ-Nash equilibrium in which all students reveal truthfully: for any μ_0 and $\epsilon > 0$, there exists $\delta > 0$ such that no student can be made more than ϵ better off by misreporting her preferences at any time in the TFM mechanism, assuming other students do not misreport.

[24] That is, if for all $\gamma > 0$ and all $\lambda \in \Lambda^{N+1}$ there exists $\omega > 0$ such that for (almost) all $\theta \in \Theta$ and all $c \in C$, $|u^\theta(c|\lambda) - u^\theta(c|\lambda')| < \gamma$ for any $\lambda' \in \Lambda$ with $||\lambda - \lambda'||_\infty < \omega$.

Finally, we show that for any μ_0 and δ there is sufficiently large T such that if the TFM mechanism converges, it does so with no student being asked her preferences more than twice, and an arbitrarily large share of students being asked only once. In other words, there are small reporting costs associated with this mechanism over canonical, one-shot mechanisms.

In what follows, we denote the outcome of the TFM mechanism (assuming the mechanism terminates) for given (μ_0, δ) as $\mu_{(\mu_0,\delta)}$, which is independent of T.

Proposition 4.

1. *Suppose that for a given economy E satisfying* ***A4*** *and a given μ_0, the TIM procedure converges to (stable) matching μ_*. Then for any $\epsilon > 0$, there exists $\delta > 0$ such that the TFM mechanism terminates in economy E and starting condition μ_0, and $\eta(\{\theta | \mu_{(\mu_0,\delta)}(\theta) = \mu_*(\theta)\}) > 1 - \epsilon$.*

2. *For any economy E, any $\epsilon > 0$, and any stopping criterion $\delta > 0$ there exists $\mu_0 \in \mathcal{A}$ such that the TFM mechanism produces an ϵ-stable matching, even when the TIM procedure does not converge.*

3. *Consider any economy E satisfying* ***A4****. Fix μ_0 and $\epsilon > 0$. Let $\Theta' \subset \Theta$ be the set of students who can profitably misreport their preferences at any step in the TFM mechanism given that (almost) no other students misreport. There exists $\delta > 0$ such that $\eta(\Theta') < \epsilon$.*

4. *For any $\epsilon > 0$ and any (μ_0, δ) for which the TFM mechanism terminates, there exists $T > 0$ such that no student θ is asked to report her preferences more than twice and the measure of students who are asked to report their preferences only once is at least $1 - \epsilon$.*

III Empirical Application: The Australian Market

The Australian higher education market uses a standardized test score and centralized clearinghouse to determine college admissions. Each year, students observe a summary statistic of the distribution of standardized test scores for the previous entering cohort at each program.

In this section we describe the details of the Australian education admissions system, and use data from this market to motivate the stylized assumptions we made in the previous section. We first discuss how students have "big fish" peer preferences over a summary statistic of student ability (AA1-AA3), and then discuss our assumptions on program entry and exit (AA4-AA6).

III.A. The Australian Tertiary Education Admissions System

We begin by describing how students apply for and attain admissions to university programs in Australia. Students apply for admission at the university-field of study (for example, Economics at University of Melbourne) level. We refer to these university-field pairs as “programs."[25]

[25] Note that tuition is regulated by the government and is not university or program specific; therefore, it should not impact applicant preferences at the program level.

Applicants receive a score known as the “Australian Tertiary Admission Rank” (ATAR) which measures the applicant’s academic rank relative to others in their age group and falls on a scale of 0-99.95. The ATAR is a central determinant of whether a student will be admitted to a program of their choice. The ATAR score is primarily determined from standardized testing, and students are not aware of their ATAR score at the onset of the application process. The ATAR score is a good predictor of academic performance during undergraduate studies (see, e.g, UAC study). Therefore, it is a proxy for student ability.

There is anecdotal evidence that students base their program selection on the ATAR scores of the previous cohort. As one student says,

> *"I was contemplating changing from Commerce/Engineering to Science/Engineering as other people who obtained a similar [ATAR] to myself were doing the Science double. Not many opt for commerce,"* (James, Baldwin, and McInnis, 1999).

Admission decisions are made in January of each year for the vast majority of students.[26] To apply for admission, students submit a rank ordered list (ROL) of up to 9 programs to a centralized admissions clearinghouse. Each program ranks applicants based (primarily) on their ATAR score. Applicants’ ATAR scores are adjusted at the program-student level with additional “bonus” points, up to 10, at the discretion of the program.[27] The clearinghouse then uses the (student-proposing) deferred acceptance algorithm to match students to programs, based on a combination of student rankings, ATAR scores + bonus points, and program capacities. In the absence of peer preferences, this algorithm makes it a weakly dominant strategy for students who prefer no more than 9 programs to their outside option to rank programs according to their *true* preferences (Haeringer and Klijn, 2009). The clearinghouse website clearly informs students of this property:

> *“First on your list should be the course you’d most like to do, followed by your second, third and fourth preferences and so on. If you’re not selected for your first preference, you’ll be considered equally with all other eligible applicants for your second preference and so on. Your chance of being selected for a course is not decreased because you placed it as a lower order preference. Similarly, you won’t be selected for a course just because you entered that course as a higher order preference.”*

This algorithm, and the resulting matching, mechanically create a minimum ATAR score

[26] Admissions in Australia take place in several rounds. We describe and analyze the process of the main round, when the majority of offers are made. There are initial rounds, where offers are made to some programs that do not admit based on the ATAR scores of students, and there are subsequent rounds for students than remain unmatched.

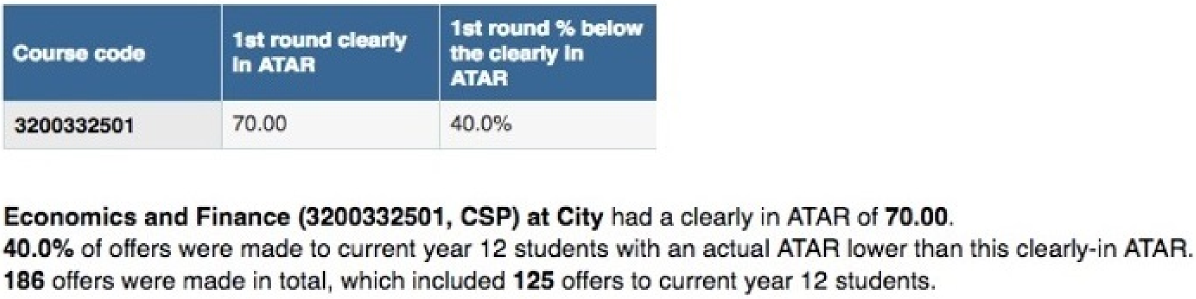
[27] Bonus points are typically awarded for reasons such as exceptional performance in school or living in a disadvantaged region.

above which students are "clearly in" (i.e. all students with ATARs above this level are admitted to the program regardless of the number of bonus points they receive if they are not admitted to a more preferred program, see Figure 1) at the program level every year.

When listing their preferences, applicants do not know the contemporaneous clearly-in ATAR statistic. They do not learn what the clearly-in ATAR statistics is until *after* the matching is generated. However, they can consult programs' clearly-in ATAR statistics from the previous year as a guide when submitting their ROLs–this information is made easily available on the clearinghouse website (see Figure 1 for a depiction of the information students are shown).[28] Going forward, we will refer to the clearly-in statistic for the cohort admitted in the previous year as the *previous year statistic (PYS)* for a particular program, and the clearly-in statistic for the current year as the *current year statistic (CYS)*. As the PYS does not necessarily equal the CYS, and because students do not know the number of bonus points they receive at each program, there is uncertainty for each applicant as to whether they will be accepted to any listed program. Across programs, roughly half of all enrolling students have ATAR scores below the CYS of their program (Bagshaw and Ting, 2016).

Applicants initially submit their ROLs before learning their own ATAR scores, but are able to costlessly change their ROLs after learning their ATAR score. Students are incentivized to submit initial ROLs early in the application process, as fees for stating initial ROLs increase over time.

Figure 1: Example of Information Provided to Applicants about a Program's Admissions Statistics in the Previous Year



Course code	1st round clearly in ATAR	1st round % below the clearly in ATAR
3200332501	70.00	40.0%

Economics and Finance (3200332501, CSP) at City had a clearly in ATAR of **70.00**.
40.0% of offers were made to current year 12 students with an actual ATAR lower than this clearly-in ATAR.
186 offers were made in total, which included **125** offers to current year 12 students.

III.B. Data

We use data from the Universities Admissions Centre (UAC), which is the centralized clearinghouse for college admissions in New South Wales and the Australian Capital Territory. Each

[28] Figure 1 shows the typical information provided to applicants during the time span covered by our data. After 2018, additional summary statistics about the previous year's ATAR distribution began to be shown.

state in Australia has a similar centralized body which processes applications to all applicants within its jurisdiction, and New South Wales is the largest state in Australia. Our data contains the universe of applications from graduating high schoolers processed by UAC from 2003 through 2016. As outlined in Section III.A. applicants initially submit their ROL without knowing their ATAR score but have the option to change their course preferences later once they become aware of their final score. For a subset of years (2010-2016) we observe applicants' ROLs at two points in time: the initial list submitted before they receive their ATAR score (which we call the pre-ROL), and the final list submitted to the clearinghouse after learning their score (which we call the post ROL). Roughly one month separates the creation of these two ROLs. We observe the post-ROL for all years in our sample (2003-2016). In addition, we observe the applicants' ATAR scores, detailed information about each program they applied to (field of study, university, and location), and the CYS of each program. We do not have information about socioeconomic background or bonus points at the application level. We are also not aware for which program each student received an offer in the end and if they chose to enroll in it. Unless otherwise specified, we use the sample of post-ROLs from 2003-2016 in our analysis.

Table 1 displays summary statistics on applicant ATAR scores, rank-ordered lists, and program PYSs. While applicants are able to list up to 9 programs, the majority rank fewer than the maximum. Going forward, we restrict our analysis to individuals who rank fewer than 9 programs on either ROL, and therefore have no incentive in the mechanism to strategically report their preferences (Haeringer and Klijn, 2009). Rows 3-6 examine the average PYS for *all* programs listed by an individual, whereas rows 7-10 focus only on the top-ranked programs. The "Avg. Pre-ATAR" PYS and score gap rows are calculated using the pre-ROL, as opposed to the final post-ROL. The "score gap" statistic is calculated using the difference between the PYS of ranked programs and the applicant's final ATAR score.

From Table 1, one can see that the top-ranked program tends to have a higher PYS than programs ranked lower on student ROLs. This PYS is also on average 6.1 points higher than the applicant's ATAR score. These statistics suggest that applicants have a general preference for higher quality schools, in so much as the PYS is a signal of school quality. It also suggests that they understand the mechanism and are not afraid of being penalized for prioritizing "reach" schools on their ROL. Table 2 formalizes this preference for quality in a regression framework. The dependent variable is the number of times a program is ranked by applicants (either first or at all) in a year, while the main regressor is the program PYS. We include field of study and year fixed effects to isolate cross sectional variation. The results show that in general, programs with higher PYSs are more likely to be ranked by students. This suggests that there is a preference for higher quality schools amongst applicants.

However, a comparison of the pre- and post-ATAR statistics suggests this demand for quality is mediated by "big fish" preferences. The score gap remains positive but narrows considerably after applicants learn their ATAR score. For example, the average gap between the top-ranked program PYS on the pre ROL and the student's eventual ATAR score is 9.9 points. After learning their ATAR score, applicants rearrange their list so that the PYS/score gap for the top ranked program is 7.9 points.

Table 1: Applicant and ROL Summary Statistics

Variable	Obs	Mean	Std. Dev.	P25	P50	P75
Student ATAR Score	471841	72.9	18.2	60	76	88
Num. of Programs Ranked	471841	7	2.3	5	8	9
Average of All Ranked Programs						
Avg. PYS	351848	79	9	72.4	78.7	85.8
Avg. Pre-ATAR PYS	205247	79.8	9	73.1	79.8	86.8
Avg. PYS/Score Gap	351848	6	13.7	-3.3	2.3	13
Avg. Pre-ATAR PYS/Score Gap	205247	7.5	14.4	-2.9	4.1	15.8
Top Ranked Programs Only						
Avg. PYS	293376	81.1	11.4	72.6	81.3	91
Avg. Pre-ATAR PYS	171860	82	11.3	74.9	82.5	91.7
Avg. PYS/Score Gap	293376	7.9	13.7	-.7	5	14.3
Avg. Pre-ATAR PYS/Score Gap	171860	9.9	14.8	0	7.1	18

This table displays summary statistics on applicant ATAR scores, rank-ordered lists, and program PYSs for all applicants in the sample. Rows 3-6 examine the average PYS for *all* programs listed by an individual, whereas rows 7-10 focus only on the top-ranked programs.

III.C. Empirical Evidence of "Big-Fish" Peer Preferences

We assume in our Section II.D..1 analysis that students have "big-fish" preferences. But do they exist in the data? Figure 2 plots the proportion of top ranked programs by score gap – i.e. the program PYS minus the ranking applicant's ATAR score. Two clear patterns emerge. The single-peaked nature of the graph suggests that students have a preference for "better" programs; the value of the horizontal axis is increasing until a positive score difference of 1. If applicants' preferences were unrelated to program quality, we would not expect the proportion of top rankings to increase monotonically with the score gap.

However, students do not want to be a "small fish" in their program of entry; the downward

Table 2: Relationship between program PYS and popularity amongst applicants

	Program ranked at all			
	(1)	(2)	(3)	(4)
PYS	0.78^{**}	0.85^{**}	0.92^{**}	1.00^{**}
	(0.28)	(0.29)	(0.31)	(0.32)
Year FE		✓		✓
Field FE			✓	✓

	Program ranked first			
	(1)	(2)	(3)	(4)
PYS	0.30^{***}	0.32^{***}	0.36^{***}	0.38^{***}
	(0.08)	(0.08)	(0.09)	(0.09)
Year FE		✓		✓
Field FE			✓	✓

This table shows the positive relationship between a program's PYS and the chance that it is included on an applicant's ROL. The dependent variable in the top panel is the number of times a program is ever ranked (any position) in a given year. The dependent variable in the bottom panel is the number of times a program is ranked first in a given year. Columns (2)-(4) include year and field of study fixed effects – this isolates cross sectional variation in the PYS across programs in a given year and field. The positive coefficients indicate that applicants generally prefer to rank programs with higher PYS's. We refer to this as a preference for program quality. Standard errors in parentheses, clustered at program level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

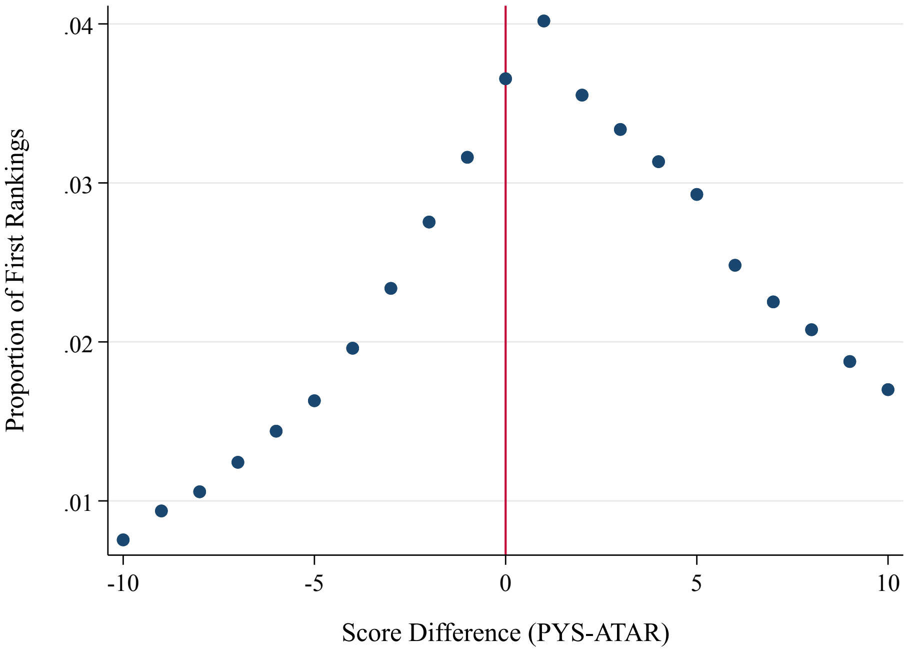
slope for score gaps greater than 1 suggests that while students are not afraid to rank "reach" schools, they become gradually less attractive as the score gap increases.

We explore this pattern of "big fish" preferences using two identification strategies below. First, we look across time at the response of applicants to changes in programs' PYS. Namely, when a program's PYS increases, does it attract applications from students with higher ATAR scores? Second, we look *within* applicant at changes in their ROL before and after learning their ATAR score. When an applicant learns their score, do they adjust their ROL to prioritize schools with similar PYSs? Both strategies show that applicants actively avoid programs with PYSs that are far above their ATAR scores.

III.C..1 Across-person analysis

When creating their ROLs, applicants have information on who was admitted to each program in the previous year (see Figure 1). How do changes in the distribution of last year's enrollees affect applicant demand for a program this year? If students prefer to attend programs where they are above or near average, then this posted information will impact their program rankings. Applicants will demote programs with PYSs that are far above their own ATAR scores. For example, all else constant, a student will be less likely to apply to a program if the PYS is 10 points, rather

Figure 2: Proportion of First-Ranked Programs, by Score Gap



This figure focuses on the top-ranked programs listed by students after they learn their ATAR score. On the x-axis is the gap between the top-ranked program's PYS and the student's ATAR score. On the y-axis is the proportion of all top-ranked programs that have that score gap. The off-center, peaked shape of the figure suggests that students understand the mechanism and have a preference for "better" programs, but at the same time do not want to be a "small fish" in their program of entry. The left side of the graph, which is increasing until a score difference of 1 point, suggests that students are more likely to rank "better", high-PYS schools, even if they are above their own score. There is no discontinuity in the figure at 0, which one might expect to occur if students misunderstood the mechanism. The downward sloping right side of the graph suggests that while students are not afraid to rank "reach" schools, they become gradually less attractive as the score gap increases.

than 5 points, above their own ATAR score.

We test for this empirically using changes in programs' PYSs across time. We rely on the fact that all students in our sample have a weakly dominant strategy to report their true preferences over programs, given their ATAR scores and the observed PYSs, to treat the submitted preferences as each student's true preferences. We estimate regressions of the form:

$$y_{c,t} = \beta PYS_{c,t} + \alpha_c + \alpha_t + \epsilon_{c,t} \tag{1}$$

where $y_{c,t}$ denotes the average applicant score, number of students who apply, percent of students who apply, or the percent of students who apply with ATAR scores higher/lower than the program PYS for program c in year t. We include year and program fixed effects (α_c and α_t, respectively) to isolate variation in the PYS that is happening within program over time. We are interested in the sign of β – when a program has a higher PYS, does it attract fewer low scoring applicants?

The results are presented in Table 3 and support our theory of "big fish" preferences. When a program's PYS increases by one point, fewer (column 2) applicants who tend to be higher scoring (columns 1) list the program in their ROL. Columns 4 and 5 test the big fish preferences discontinuously – the dependent variable splits the sample into individuals with scores either above or below the PYS of program s, and quantifies the percentage who have listed program c. As shown in column 5, the big fish effect is driven primarily by those with scores *below* the PYS. They become less likely to rank the program on their final ROL.

Table 3: Across Time Applicant Response to Program PYS

	(1)	(2)	(3)	(4)	(5)
	Avg. Applicant Score	# of Applicants	% of Applicants	% of Applicants Higher Score	% of Applicants Lower Score
Past Year Statistic	0.344***	-2.709***	-0.008***	-0.003	-0.015***
	(0.015)	(0.267)	(0.001)	(0.001)	(0.001)
Observations	14,850	14,850	14,850	14,850	14,850

This table shows the estimated β coefficients of equation (1) were $y_{c,t}$ is the average applicant score, number of students who apply, percent of students who apply, or the percent of students who apply with ATAR scores higher/lower than the program PYS for program c in year t. Standard errors in parentheses, clustered at program level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Interpreting these results as evidence of peer preferences requires several identifying assumptions. First, it assumes that applicants form preferences using programs' PYS – in other words, they believe that the previous year statistic is equivalent to the current year statistic. To test this assumption, we re-run the specification while including values of a program's CYS and lagged PYS in Table A.2. If students naively assume that the PYS is equivalent to the CYS, then coefficients on CYS should be near zero, aside from the mechanical positive effect of having a higher scoring population apply.[29] We find that the coefficients on the CYS are generally an order of magnitude smaller than on the PYS, and statistically less significant.

We also test whether the PYS coefficient is simply picking up trends in program quality – we would observe a positive correlation between PYS and average applicant score if a specific program is becoming "better" over time. However, lagged values of the entry statistic (from two or three years before the current year) have little predictive power when included in our across-time regression.

A third concern is that applicants do not use the PYS to learn about the skill distribution of peers, but rather as an indication of their "fit" with a particular program. They may be uninformed

[29] For example, if students with ATAR scores higher than the PYS apply in greater numbers, this will mechanically lead to a higher CYS. A small, positive coefficient on the CYS reflects this mechanism, and does not necessarily show that students anticipate the CYS.

about a program and interpret the PYS as a sign, for example, of how difficult or prestigious the program is. Their choices may be influenced by these updates to their program-specific information rather than the peer preference mechanism. To rule out this channel, we interact the PYS with program age. This test leverages the fact that applicants likely have more information about long-standing programs. Changes in the PYS provide relatively less information about older, established programs. Under the "program information" hypothesis, the effect of the PYS on applicant demand should dissipate for older programs. In Table A.3 we show that effects are generally statistically similar for old and young programs, suggesting that students are not learning about other characteristics of programs through the PYS.

Using application data to measure peer preferences also requires that applicants *understand* the deferred acceptance mechanism. The clearinghouse website prominently instructs students to rank their preferences truthfully, emphasizing that the algorithm is strategy proof. However, it is possible that applicants misunderstand the assignment process and two types of misunderstanding appear plausible and could result in patterns consistent with big fish peer preferences. First, students may believe that being rejected by a program high on their preference list will make it less likely to be matched with a subsequent program, and second, students may believe that their probability of admission is 0 if their ATAR score falls below the PYS. Both of these forms of misunderstanding could result in a student failing to list a preferred program with a PYS greater than her ATAR score.

We do not find evidence for either of these possibilities. Over 75% of applicants rank at least one "reach school" (defined as having a PYS that is higher than the applicant's ATAR score) first. Moreover, if a substantial subset of students had one of these forms of misunderstanding, we would anticipate a discontinuous drop in the share of students ranking a program first that has wth a PYS just above their ATAR score versus just below their ATAR score. Figure 2 plots the proportion of top-ranked programs by the difference between the program's PYS and the student's ATAR score. There is no discontinuity in the figure at 0 along the horizontal axis. Indeed, the modal score difference is +1, suggesting that a program with a PYS just above a student's own ATAR score is most likely to be ranked first.

Finally, an important concern in attempting to glean preference information from ROLs in strategy-proof mechanisms is the possibility that students make mistakes when reporting (Chen and Sönmez, 2002; Chen and Pereyra, 2019; Sóvágó and Shorrer, 2018; Hassidim, Romm, and Shorrer, 2020). However, as such mistakes are often payoff irrelevant in higher-education markets; in a similar market, Artemov, Che, and He (2020) argue that students are well aware of the rules of the matching mechanism, and that “mistakes" generally occur when students fail to list unattainable options at the top of their ROLs. Artemov, Che, and He (2020) upper bound the share of applicants making observable, payoff relevant mistakes in their own study at .72%, in Hassidim, Romm, and Shorrer (2020) at 1.5%, and in Sóvágó and Shorrer (2018) at 1.6%. As we have said, 75% of students in our sample rank a reach program first on their ROL, indicating that such strategic "mistakes" may be rare. Nevertheless, we conduct the following identification strategy to mitigate any such concerns.

III.C..2 Within-person

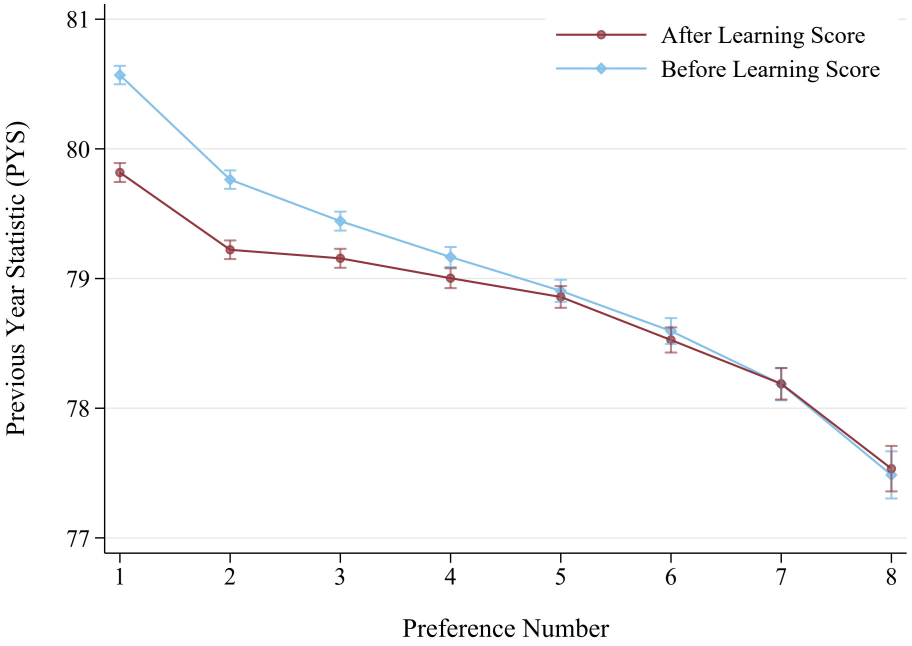
In addition to using changes in program selectivity over time, we also measure how applicants respond to new information about their own ability. We observe applicants' ROLs at two points in time: both before and after they learn their final ATAR score. Students are incentivized monetarily to complete their ROL early, before receiving their final ATAR score. However, they can update their ROL after learning their score. To draw inferences about student preferences from ROLs, we additionally assume that pre-ROLs reflect true preferences, taking into account expected ATAR scores. We discuss this assumption later in this section.

We find that students frequently update their rankings after learning their scores, and that these changes meaningfully affect their final matchings. Applicants tend to rearrange their ROL to deprioritize programs with PYSs far above their realized ATAR score. This effect is especially prevalent for lower scoring applicants.

Figure 3 plots the average program PYS by position on the rank-ordered list, both before and after ATAR scores become known. Preference number 1, on the x-axis, refers to the top-ranked program. In general, individuals rank programs with higher PYSs earlier on their list. They also update their lists after learning their score by listing less competitive programs, in particular at the top of their list. The PYS gradient from top to bottom ranking remains downward sloping, but becomes less steep.

We next split the sample into high and low scoring students (see Figure 4) to test whether these ROL changes are suggestive of big-fish preferences. Changes in rankings correlate with an applicant's ATAR score. Applicants who have low ATAR scores (near the 25th percentile) make the largest adjustments in rankings. On average, they replace high-PYS programs near the top of their lists with programs that have lower PYSs, closer to their own ATAR score. On the top end of the distribution, high scoring applicants make a small, but statistically significant, shift towards higher PYS programs. While moving in the opposite direction, this again closes the "gap" between applicant ATAR score and program PYS. In general, after the release of ATAR scores, we see applicants adjusting their lists to prioritize programs with PYSs that are closer to their own scores. This pattern suggests that not all students have perfect knowledge about what their final ATAR score will be when making their initial lists, and it is consistent with a desire to avoid programs where peers have significantly higher test scores.

Figure 3: Average Listed Program PYS by Rank Order, before and after Score Revelation



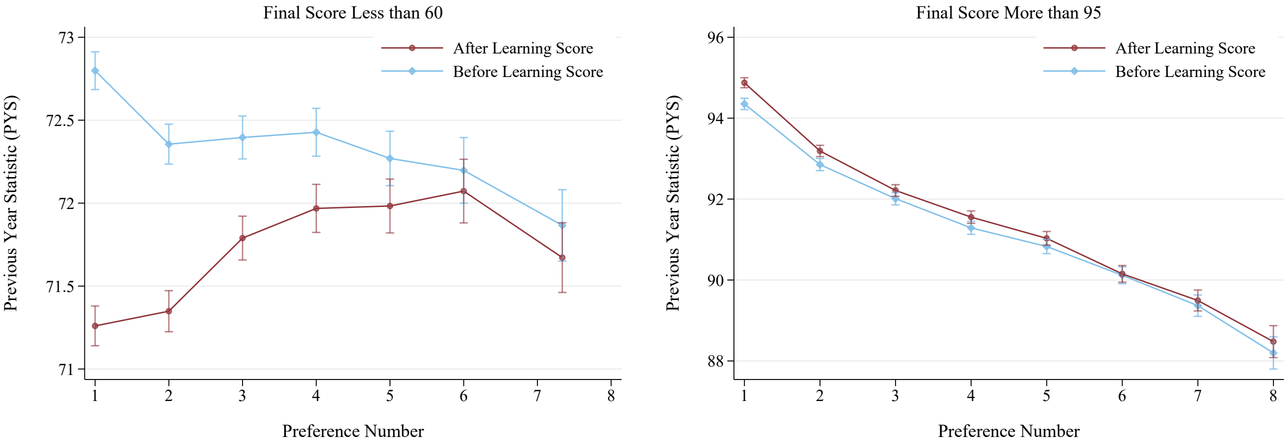
This figure plots the average PYS of programs listed by applicants before and after they learn their own score. The x-axis denotes the position of a program on the rank ordered list. This figure restricts to individuals who ranked fewer than the maximum number of programs. On average students rank 1) rank higher PYS programs at the top and 2) modify their preferences with lower PYS programs after they learn their score. We use the pre- and post-ROL sample from 2010-2016.

We investigate *how* students adjust their ROLs after learning their ATAR scores to more fully illuminate the effect of peer preferences. Students can adjust their pre ROLs in three ways. They can *add* a program, they can *remove* a program, or they can *switch* the relative rankings of two programs. A switch is defined as an instance where program c is ranked higher than program c' on the pre ROL, both c and c' are on the post ROL, and c' is ranked above c on the post ROL. In this case, c' is *promoted* and c is *demoted*.

Note that switches should not occur even if students play a weakly dominated strategy in which they omit programs from their ROLs that they have a a very low probability of gaining admission to (Artemov, Che, and He, 2020; Fack, Grenet, and He, 2019). Because the matching mechanism is strategy proof, switches cannot be easily explained as occurring purely because of a student's assessed probability of admission at a program. Therefore, the occurrence of switches strongly suggests that students wish to attend different programs after observing their test scores.

Appendix Table A.1 shows 41% of students do not submit the same pre ROL and post ROL. On average, each student makes 1.77 adjustments, which corresponds to 27% of the pre ROL. Of the adjustments made, switches are the most common.

Figure 4: Average Listed Program PYS by Rank Order, before and after Score Revelation for low vs. high scoring students)



These figures plot the average PYS of programs listed by applicants before and after they learn their own score. We use the pre- and post-ROL sample from 2010-2016. The x-axis denotes the position of a program on the rank ordered list. Both figures restrict to individuals who ranked fewer than the maximum number of programs. The left figure restricts to students who receive ATAR scores strictly below 60 (low scoring students), while the right figure restricts to high scoring students. On average, low scoring students modify their rankings "downwards" with lower PYS programs after they learn their score. For high scoring students, there does not appear to be large changes on average following score revelation.

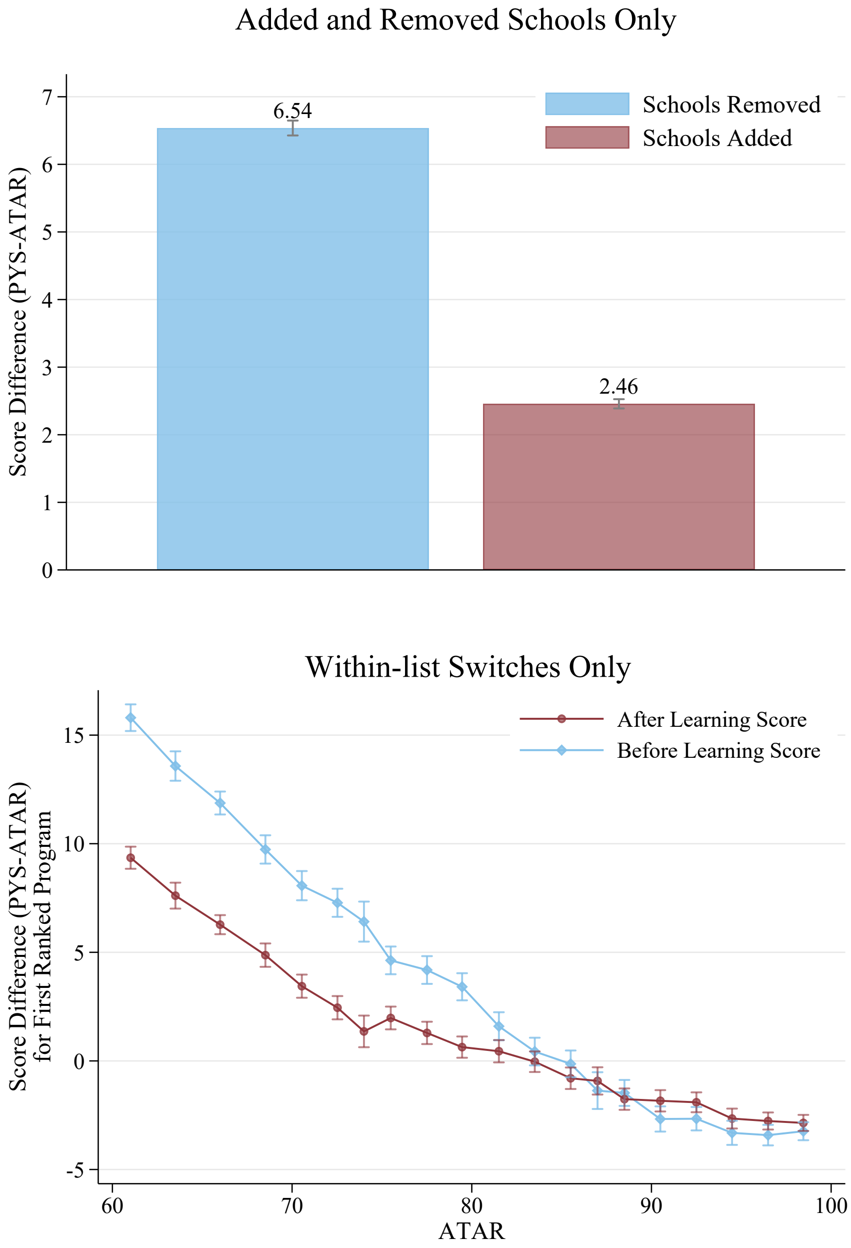
Adjustments that students make–additions, removals, and switches–result in a smaller PYS/Atar score gap –i.e. the difference between the program PYS and the applicant's ATAR score.[30] The top panel of Figure 5 plots the PYS-score gap for programs that are added or removed from ROLs. This extensive margin maneuver reduces the average PYS/ATAR score gap from 6.54 to 2.46 points. Similarly, when we focus on programs that are switched (see the bottom panel of Figure 5), the absolute value of the PYS/ATAR score gap for top-ranked programs shrinks at both the top and bottom ends of the ATAR distribution; students with ATAR scores below 85 generally promote a lower PYS program to the first choice, while those with ATAR scores above 85 generally promote a higher PYS program. In line with big fish preferences, this change in sign occurs for ATAR scores such that students are, on average, ranking programs with PYS<ATAR.

We quantitatively analyze these effects through linear regressions. Specifically, we regress the PYS/ATAR score gap on indicators for whether that program was removed, added, or promoted–i.e. switched "upward"–within the applicant's ROL after they learn their ATAR score.[31] We run

[30] All of our within-applicant empirical results are robust to restricting to students who make "only switches" and/or make "only additions/removals".

[31] The sample for this table is all students who rank at most 8 programs in both their pre and post ROLs. The variables "remove" and "add" indicate when a program is removed (or added, respectively) from an individual's ROL after learning their ATAR score. We classify a program as "promoted" if it appears on both the pre and post ROL and is in a relatively higher spot on the post-list than on the pre-list, ignoring all other adds and drops. To define promotion, we work with the following inversion algorithm:

Figure 5: Average Listed Program PYS before and after Score Revelation, Restricting to Added/Removed Programs (top panel) and Switched programs (first-ranking only, bottom panel)



This figure plots the average gap between the admissions PYS for programs listed by applicants and the applicant's ATAR score. The top panel looks only at programs that were either added or removed from the applicant's list after learning their score – in general programs that are added after learning the ATAR have significantly lower PYSs than those that are removed. The bottom panel looks at the gap for top ranked programs that are switched elsewhere in the list after the applicant learns their score. Again, lower scoring applicants rearrange their lists to prioritize lower PYS programs. We use the pre- and post-ROL sample from 2010-2016.

- Keep only programs that are on both the pre- and post-list (i.e. remove all adds and drops from both lists), and call these the redacted pre- and post-lists, respectively.

the following regressions

$$y_{c,t,i} = \beta(PYS_{c,t} - ATAR_i) + \alpha_c + X_i + \epsilon_{c,t,i} \tag{2}$$

where c represents the program, t the year, and i the student. $PYS_{c,t} - ATAR_i$ represents student i's score gap at program c in year t, α_c represents a program fixed effect, and X_i represents a vector of pre ROL characteristics for student i (including the identities of the top ranked, second highest ranked, and third highest ranked programs, the average PYS across all programs, and the number of programs included on the pre ROL). The dependent variables studied are whether the program c is removed from the pre ROL, added to the post ROL, or promoted in the post ROL.

Table 4 displays the results from these regressions. Programs that are added or promoted within the ROL have PYSs that are systematically lower than those that are removed, and closer to the applicant's ATAR score. Programs that are removed have a larger score gap – i.e. they are more of a "reach" school. These effects are consistent with big fish prefernces. These effects persist with an array of fixed effects. In Columns 3-7 we attempt to compare the behavior of applicants who construct very similar ROLs *before* learning their ATAR score by including fixed effects for ranking the same programs or programs with the same average PYS. For example Column 7 provides evidence of how the adjustments made by students whose pre ROLs have the exact same three programs ranked in the top three spots, have similar average PYS of programs ranked on the pre ROL, and the same number of programs ranked in the pre ROL changes given different ATAR scores received. Under our identifying assumption that pre ROLs represent true preferences given expected ATAR scores, this provides evidence on how students with similar underlying preferences heterogeneously adjust their ROLs following different "shocks." We observe the existence of big-fish preferences even amongst these groups of applicants who have similar pre-ATAR preferences but different ATAR realizations.

In order to interpret these within-person patterns as indicative of peer preferences, we must make specific assumptions. We assume that each applicant's pre ROL reflects the true ranking of listed programs, given expected ATAR score and the assumption that PYS=CYS. In other words, applicants truthfully list their preferences, rather than an arbitrary set of programs, before learning their ATAR score. We require a relatively weaker assumption, that the *relative* rankings of any two programs on the pre ROL reflect true preferences–given expected ATAR score and the assumption that PYS=CYS–when we focus only on programs that are promoted or demoted within

- A program is promoted if it is ranked in a higher spot on the redacted post-list than on the redacted pre-list.

Note that any switch results in one program being promoted and one program being demoted. As a result, we do not include "demote" in this regression.

Table 4: Adjustments to ROL by PYS/ATAR Score Gap

	(1) Score Gap	(2) Score Gap	(3) Score Gap	(4) Score Gap	(5) Score Gap	(6) Score Gap	(7) Score Gap
Remove	0.648***	0.773***	1.301***	1.572***	1.658***	1.911***	1.949***
	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Add	-3.957***	-2.660***	-2.793***	-2.061***	-1.485***	-1.601***	-1.401***
	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Promote	-0.579***	-0.756***	-1.130***	-1.328***	-1.358***	-0.263***	-0.353***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Constant	6.659***	6.560***	6.632***	6.586***	6.526***	6.227***	6.227***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Program FE		✓					
Top ranked program FE			✓				
Top two ranked programs FE				✓			
Top three ranked programs FE					✓	✓	✓
Avg. ROL PYS FE						✓	✓
ROL length FE							✓
Observations	579,987	579,958	579,987	579,987	579,987	578,552	578,552

Dependent variable is the score gap to PYS in all regressions. Column (3) includes student fixed effects, column (4) includes program fixed effects, column (5) includes a fixed effect for the top ranked program in the pre-list, column (6) includes a fixed effect for the top two ranked programs in the pre-list, column (7) includes a fixed effect for the top three ranked programs in the pre-list, column (8) adds a fixed effect for the average PYS of the pre-list, and column (9) adds a fixed effect for the total number of schools ranked on the pre-list. We use the pre- and post-ROL sample from 2010-2016. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ROL.

There are several encouraging facts in the data that support this assumption. First, students are incentivized to submit preferences early through lower application fees, before learning their ATAR score. The overwhelming majority (99.5%) of applicants do so. Second, there is a high correlation between applicants' pre- and post- ROLs, as seen in Appendix Table A.1. This suggests that students do not rank an arbitrary list which they need to change completely. Finally, changes (especially switches) to an applicant's ROL are predicted by the difference between the realization of their ATAR score and PYS of a program. If the initial ROL was arbitrary, then we would expect no correlation between the score gap and the number of ROL changes.

The evidence does not support that adjustments to the pre ROL are due to preference changes over time unrelated to peer preferences. From a timing standpoint, only one month separates our observation of the pre and post ROLs. Moreover, the fact that adjustments to students ROLs are predicted by their realized test scores does not support this alternative hypothesis. Specifically, for preference changes over time to rationalize our findings, it would have to be that programs ranked closer to a student's eventual ATAR score are systematically receiving a positive "shock" relative to other programs.

III.C..3 How important are peer preferences?

How much do peer preferences affect the final matching? While the adjustment of ROLs suggests that applicants prefer not to engage with peers of significantly higher ability, is the effect of these preferences large at the market level? Do individuals have a lower probability of attending a program with a PYS far above their own score?

The data suggests that they do. Figure 6 plots applicants' average probability of acceptance to each program on their final ROL using the pre-ATAR ROL (in blue) and the post-ATAR ROL (in red). On average, a student has a 60% chance of being matched to the top ranked program on her final ROL, but the probability of matching with this program would drop to 40% if she did not update her pre-ATAR ROL.[32]

We can calculate the approximate[33] share of students whose final matchings are changed from the counterfactual world in which that student has instead submitted her pre-ATAR ROL as her final ROL as:

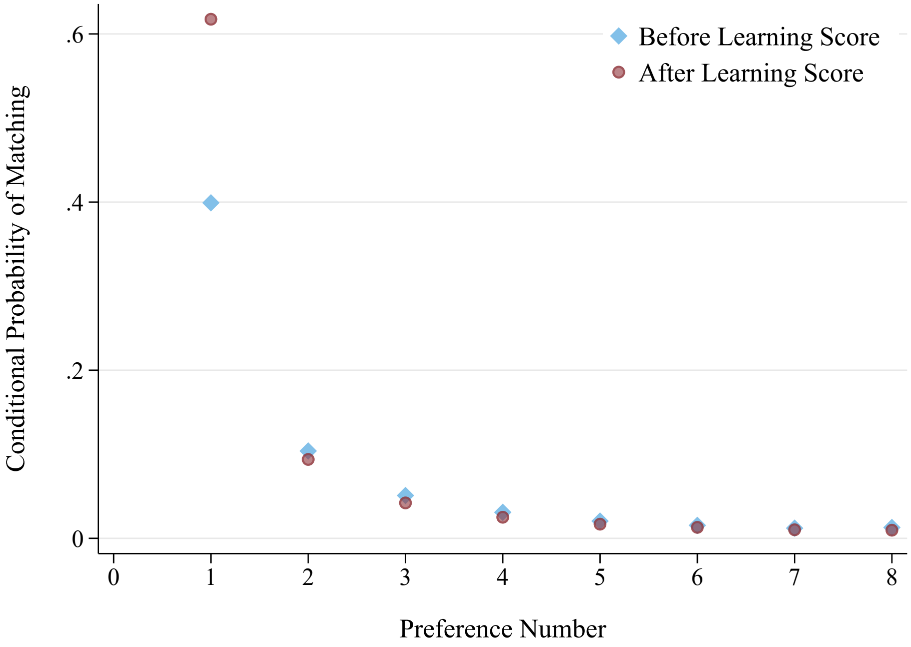
$$\begin{aligned} \approx \sum_{j=1}^{8} \quad & \text{Pr(Matched to preference number } j \text{ in post ROL)} - \\ & \text{Pr(Matched to preference number } j \text{ in post ROL } \textit{if instead submitted pre ROL}) \end{aligned}$$

In Figure 6, the approximate share of students who would receive a different final matching in the counterfactual world is the sum of the absolute value of the difference between the red and blue dots for each preference number. The figure shows that these changes are meaningful – the final ROL increases the probability of getting one's first choice program by 22%, and of receiving a different final matching by 25%. Under our identifying assumption that the post-ATAR ROL represents true student preferences, each student is worse off in the counterfactual world. Our identifying assumption that students submit the pre-ATAR ROL reflecting their *expected* peer preferences also suggests this calculation is a *conservative* estimate of peer preferences; students' expectations, and hence their pre-ATAR ROL, will already reflect some of their peer preferences.

[32] For robustness, we repeat this exercise but vary how we calculate admissions probabilities. Since admissions are calculated based on whether the sum of a student's ATAR points and program-specific bonus points is greater than the program's CYS, this exercise amounts to simulating various distributions of bonus points. The Figure 6 is created assuming that the number of bonus points awarded to each student at each program is a uniform random variable with support $\{0, 1, ..., 10\}$, and is independent across students and programs. We also run an optimistic scenario in which the number of bonus points awarded to each student at each program is 10 and a pessimistic scenario in which the number of bonus points awarded to each student at each program is 0. A similar result holds at either extreme.

[33] The reason this calculation is approximate is that the length of each student's pre-ATAR and post-ATAR ROL is not necessarily equal. However, as seen in our summary statistics table, the length of pre- and post- ROLs are close for most students.

Figure 6: Conditional Probability of Matching with Programs before and after learning Score, by Rank Order



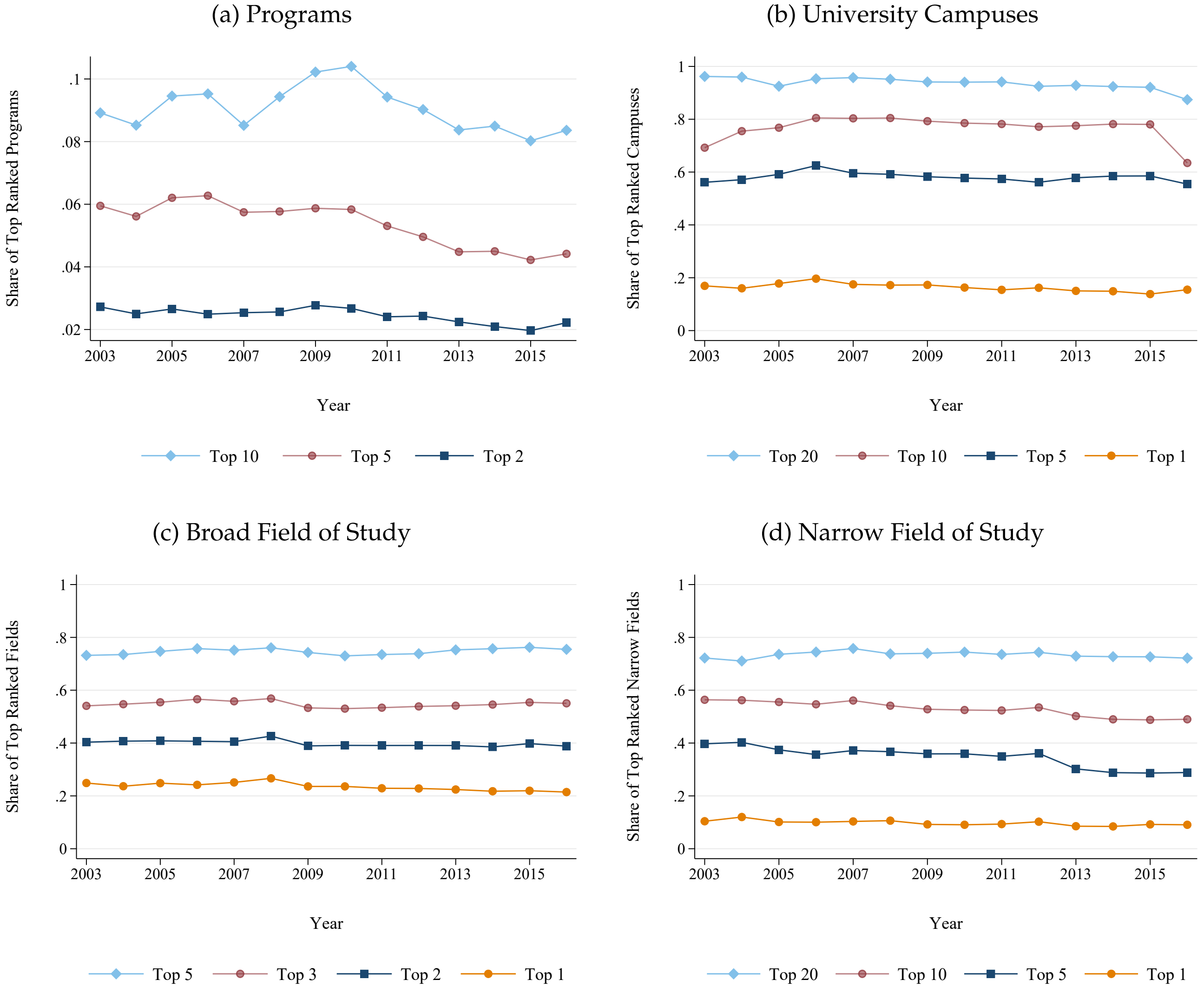
Notes: we calculate an applicant's probability of matching to each program on their ROL. We do this calculation based on both their rankings before and after learning their ATAR score. For each student-program pair, we independently (across both students and programs) assign a number of bonus points, assuming a uniform random variable with support $\{0, 1, ..., 10\}$. A student is matched to a program if it is the highest program on her ROL such that her ATAR score plus assigned bonus points exceeds the CYS of the program. We use the pre- and post-ROL sample from 2010-2016.

III.D. Empirical Evidence of Changes in the Market

In Section II.D..1 we assume in our analysis that the set of programs changes from year to year, but that other factors are not changing simultaneously within the Australian market – for example the popularity of certain fields of study, the average ability of applicants, or the popularity of certain universities. We test these assumptions by looking at the distribution of additional variables (other than the PYS) over time. Figure 7 shows that aggregate student preferences for field of study and university campus appear to remain relatively constant over time, while aggregate student preferences over programs vary more from year to year.

When we examine the distribution of programs over time, we find that there is significant entry and exit in the set of offered programs. Moreover, our analysis in Section II.D..1 assumes that entry and exit of programs occurs for programs that have a low PYS. Figure 8 shows the somewhat bimodal distribution of program "age" – while some programs are very established and exist for 14 or more years, the majority only exist forfewer than 4 years. In addition, we plot the difference in PYS for each age cohort relative to the youngest programs. Programs that enter

Figure 7: Aggregate student preferences over programs, campuses, and fields


(a) Programs

(b) University Campuses

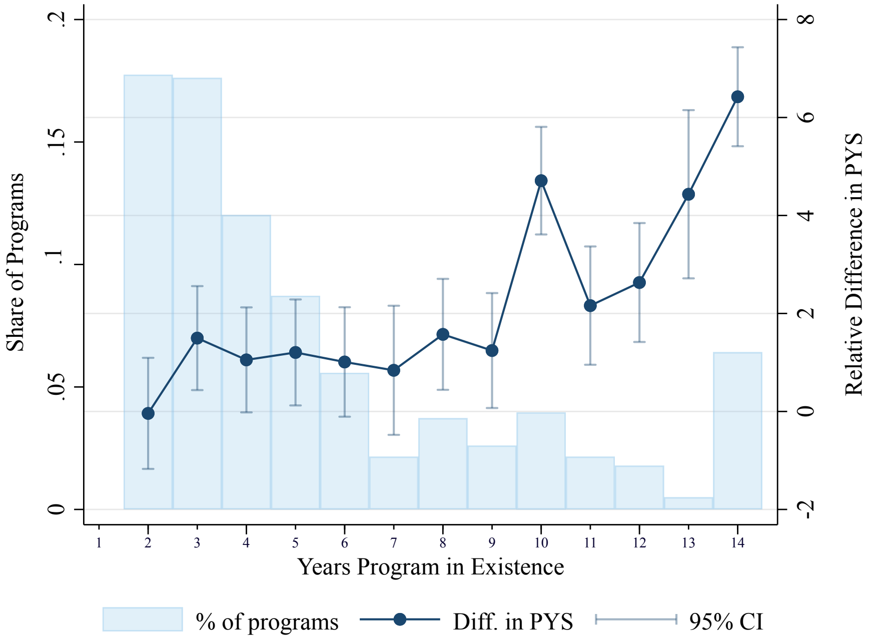
(c) Broad Field of Study

(d) Narrow Field of Study

This figure shows that applicant preferences for program, campus, and field of study are broadly stable across cohorts. We "fix" a group of campuses or programs based on overall popularity, and then show that this popularity ranking is stable year to year. To create the campus graph, for example, we use the entire panel dataset of applications. We count how many times each campus was ranked either first or second on an applicant's list. This provides an "overall" measure of popularity that can be used to rank the campuses. We then define groups containing the X most popular overall campuses (each line on the graph is a different size of X). We plot the market share (as defined by how many times it was ranked first or second on an applicant's list) for that group of X campuses in each year. The resulting graphs show that a small group of campuses and fields *consistently* remain the first or second choice for the majority of students. For example, the yellow line, which refers to the number 1 most popular overall campus, consistently receives the top ranking for about 20% of applicants each year. The navy blue line, which refers to the group of top 5 most popular overall campuses, consistently make up about 60% of top rankings each year.

and exit frequently have lower statistics than incumbents. In the following section, we investigate the impact of entry and exit of less popular programs on stability.

Figure 8: Difference in PYS by length of program existence



This figure plots the estimated difference in PYSs for programs based on how long they exist in our data, relative to programs that exist for only one year. These estimates control for the initial calendar year in which the program enters, and the field of study. Programs that exist for 14 years, for example, have on average a PYS that is almost 7 points above programs that only exist for 1 year. The upward sloping pattern to the point estimates supports the hypothesis that programs entering and exiting generally have lower PYSs than the programs that are more established. Underneath the point estimates, we also overlay a histogram that shows the distribution of years of program existence. There is somewhat of a bimodal distribution – most programs have rapid entry and exit (i.e. they exist for only 1-3 years), whereas another significant portion exists for 14 years.

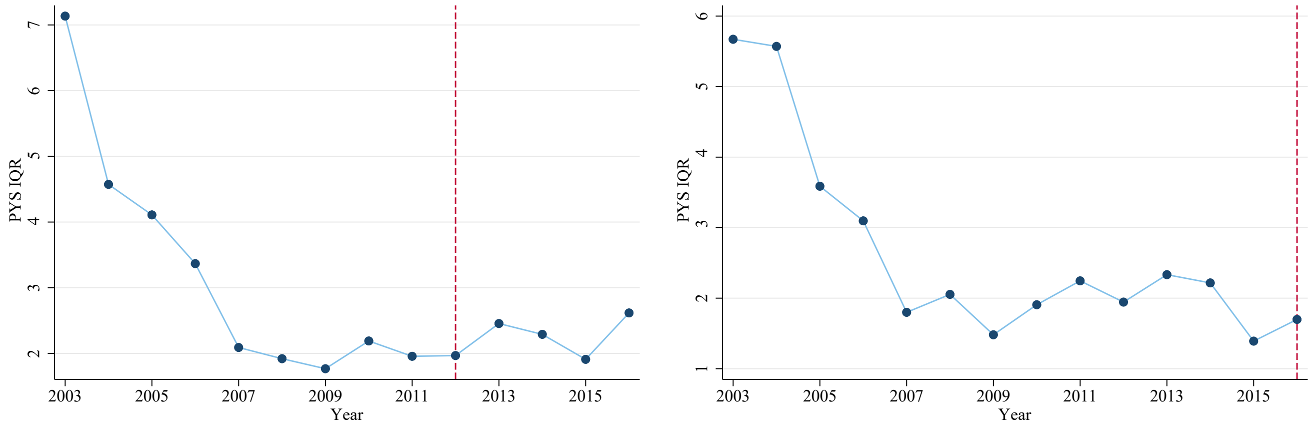
IV Does ignoring peer preferences generate instability in Australia? Empirical Evidence

First, we test the stability of the Australian market using data on the evolution of program PYSs over time. We show that when a program initially enters the market, there is substantial volatility in its PYS year-to-year. However, this volatility decreases with time, and almost entirely stabilizes by year 12. There is instability in the market, however, due to frequent entry and exit of low-popularity programs. We show theoretically that students matched to low-popularity programs will in general not receive their stable partner even in the long run, while high-scoring students who enroll in more popular programs will receive their stable partner in the long run. Empirically, we show that volatility in PYS is correlated with higher attrition, consistent with the notion that these students are more likely to be involved in blocking pairs. We also show that indigenous, low SES, rural, and students with disabilities are less likely to be matched to their stable-matching partner.

Theorem 2 and Corollary 2 state that a market delivers a (approximate) stable matching in the long run if and only if the PYS of each program converges over time.

Figure 9 provides evidence of this convergence. Panel 1 plots the interquartile range of the PYS by year of programs with PYSs between 65 and 75 in 2012, while Panel 2 plots the interquartile range of the PYS by year of programs with PYSs between 65 and 75 in 2016. In both panels, there is smaller year-by-year change in the years immediately preceding the base year, than in the initial years. Moreover, Panel 1 suggests that this is not merely due to mean reversion; in the years following 2012, the mean of program PYSs remains nearly constant, and is less variable than in the years immediately preceding 2012. If these plots were driven largely by reversion to the mean of the PYSs in the base year, we would not expect the PYSs in years 2013-2016 to remain nearly constant.

Figure 9: Convergence test for programs with ultimately similar PYSs



The top figure groups together programs that have a similar PYS (within a 10 point band of 70) in 2012. It then follows the group's distribution of PYSs (as measured by the interquartile range) both forward and backwards in time. It shows that programs with similar PYSs in 2012 have converged from a more disperse distribution over time, and do not appear to diverge after 2012. The bottom figure repeats the same exercise, instead grouping together programs with a similar PYS in 2016. In the appendix we plot a similar set of graphs (see Figure A.1) that show the progression of the groups' mean PYSs over time. They display a very similar pattern, in which the average PYS converges over time.

Panel 1 of Figure 10 plots the average absolute change in program PYS by age of program, where programs experience larger changes earlier in time. Our data allow us to track programs for a maximum of 14 years, and we observe that programs initially have a year-to-year change of nearly 2.5 points, while programs in years 12-14 have an average point estimate change in their PYS of half a point.

Panel 2 controls for entry and exit of programs into our data by grouping programs that by number of years they enter our data, and recreating the plot in Panel 1 for each group. Across all groups, we observe a similar decreasing trend in the absolute change in PYS over time, that falls under 1 point as programs age beyond 10 years (4 of 5 of the groups in our data for at least 10 years have all point estimates beyond year 10 less than 1 point.) Nevertheless, this suggests that while the PYSs of individual programs are converging over time, entry and exit of programs could cause instability in the market if programs are not present in the market long enough for their PYSs to reach (near) steady state.

IV...1 Stability, Attrition and Student Demographics

The previous section finds that students matched to high-popularity programs are less likely to be involved in a blocking pair. Moreover, only students matched to low-popularity programs will be involved in a negative utility blocking pair.

To discuss the impact of this potential instability, we first focus on one related outcome: attrition, which occurs when a student neither graduates from the degree program, nor returns in the following year. Whenever a blocking pair is consummated, attrition occurs, and therefore, we expect attrition to be higher at programs with students who have more blocking pairs, especially negative utility blocking pairs–as only students who have negative utility blocking pairs can prefer being unmatched to remaining at their current program. For privacy reasons, we do not observe attrition at the individual level, but we instead merge in the attrition rate at the university-year level. As we do not observe individual blocking pairs, we rely on results of Theorem 3 and Remark 4: students at programs with larger absolute changes in the PYS are more likely to be in blocking pairs, and only students at programs where the CYS<PYS are members of negative utility blocking pairs. Table 5 shows that, at the program level, higher yearly changes in the PYS are correlated with higher attrition rates. When we restrict to programs-years with CYS<PYS (i.e. the threshold is falling over time), the correlation is even stronger. This relationship is not driven by yearly trends or field-specific patterns (it is robust to year and field fixed effects). Is is also not driven by program age or size.

Moreover, we find that the consequences of failing to explicitly design the market to incorporate peer preferences is borne in stability terms by students from less advantaged demographic backgrounds. We merge in data on gender, ethnicity, and socioeconomic status at the university-year level.[34] We test for a significant relationship between yearly absolute changes in PYS and the share of low SES students in Table 6. This relationship is positive and significant — a one percentage point increase in share of low SES students at a university is associated with an increase in the yearly PYS change measure of .012 points. The average yearly absolute change in

[34] Due to student privacy concerns, we are not able to merge demographic characteristics at the individual level.

Table 5: Relationship Between |CYS-PYS| and Attrition Rate

	All Observations (N = *14,795*)				
Attrition Rate	0.011* (0.005)	0.009 (0.005)	0.010* (0.005)	0.020*** (0.006)	0.004 (0.015)
	Only program-years with CYS-PYS < 0 (N = *4,226*)				
Attrition Rate	0.158*** (0.011)	0.165*** (0.012)	0.152*** (0.012)	0.138*** (0.014)	0.183*** (0.030)
Year FE	✓	✓	✓	✓	✓
Field FE		✓	✓	✓	✓
Course Age FE			✓	✓	✓
School Size FE				✓	✓
Field Shares FE					✓

This table tests for the relationship between the attrition rate of a given program and its year to year change in PYS. We find that, even with a host of controls and fixed effects, programs with more volatility in their yearly admissions statistic also have higher student attrition rates. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

PYS is .35, so from a percentage view this is a 3.4% increase off the mean. This relationship is robust to controls for year, field of study, program age, and school size. The relationship is several times the magnitude when we restrict to program-years with CYS<PYS, i.e. those that admit negative utility blocking pairs. We find a similar pattern when looking at the share of minority, disabled, and rural-based students across schools with more or less volatile PYS (see Figure A.2 in the appendix).

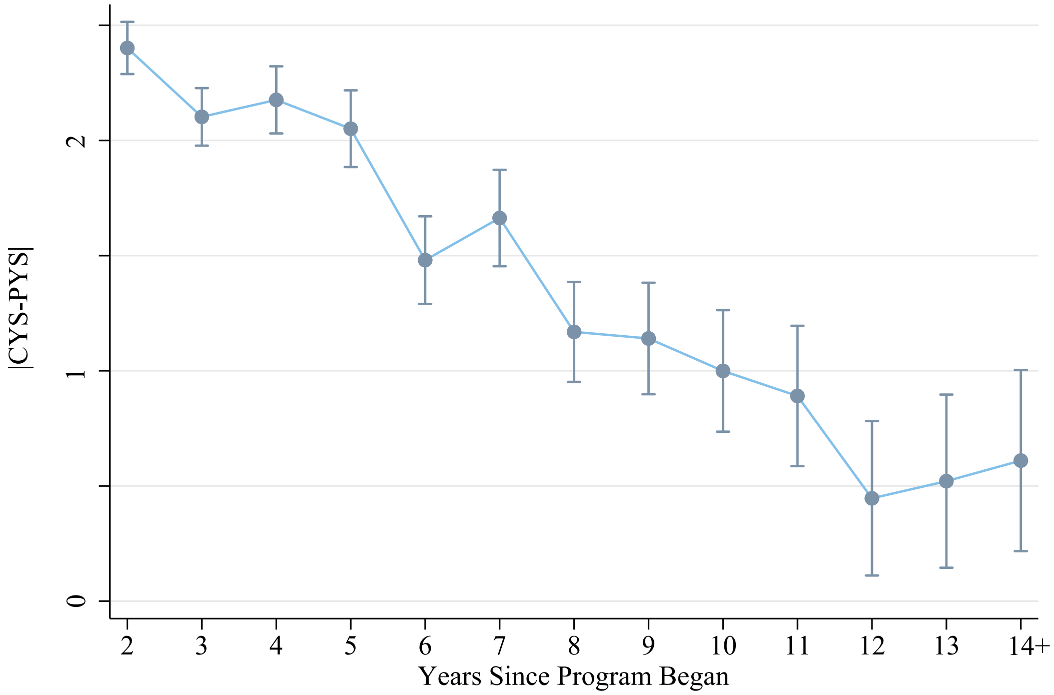
These results suggest that programs with less stable statistics are more likely to serve a lower SES population, and are subject to higher attrition rates. While these results do not themselves convincingly show causality in either direction, they results do show that the population most impacted by non-convergence has a lower socioeconomic status, and those who are less likely to complete their studies at their initial program.

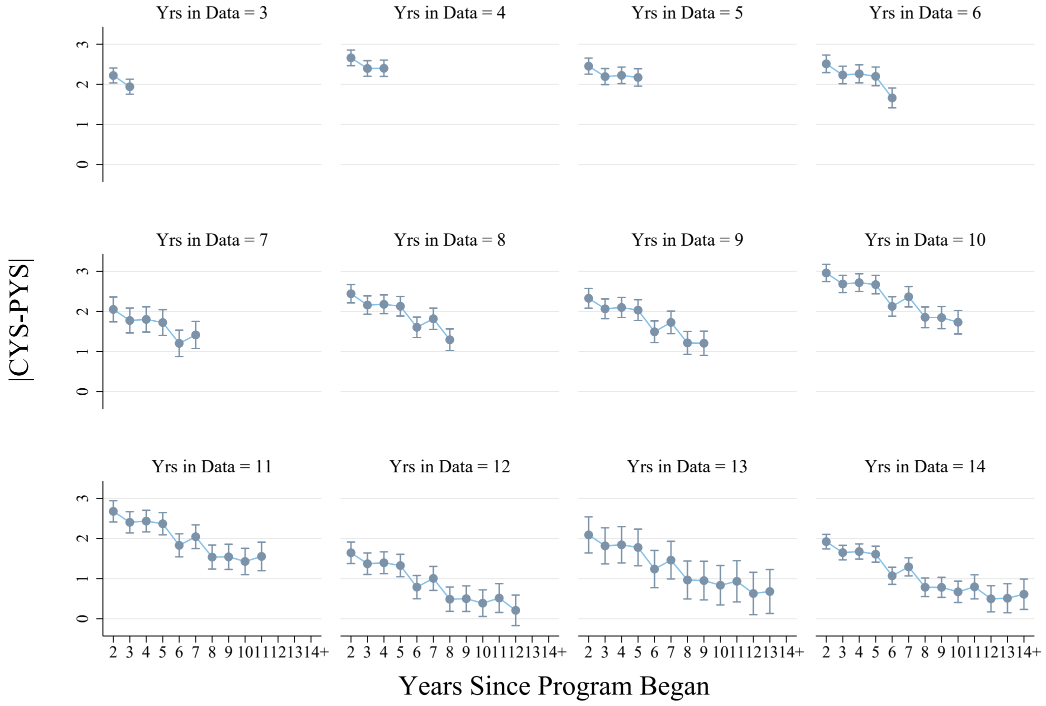
Table 6: Relationship Between |CYS-PYS| and Share of Low SES Students

	All Observations (*N* = *14,795*)				
Share Low SES	0.012***	0.010**	0.012***	0.013***	0.027**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.009)
	Only program-years with CYS-PYS < 0 (*N* = *4,226*)				
Share Low SES	0.064***	0.061***	0.059***	0.055***	0.162***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.020)
Year FE	✓	✓	✓	✓	✓
Field FE		✓	✓	✓	✓
Course Age FE			✓	✓	✓
School Size FE				✓	✓
Field Shares FE					✓

This table tests for the relationship between the share of students with a low socioeconomic background of a given program and its year to year change in PYS. We find that, even with a host of controls and fixed effects, programs with more volatility in their yearly admissions statistic also have higher share of low SES students. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 10: Year-to-Year Variation in PYS Within Program, Over Time





These figures demonstrate that there seems to be convergence over time in program PYSs. For each program we plot $\Delta_{c,t} := |CYS_{c,t} - PYS_{c,t}|$. The graphs plot this measure of PYS "instability" against the number of years the program has existed. They show that the PYS seems to converge with time (top figure), and that this panel is not driven by the entry or exit of programs into the sample (bottom figure)

V Conclusion

How important is it that a matching market allows agents to fully express their preferences? We study this question in a market where students have preferences over their peers but cannot express these in the matching mechanism. We show that a dynamic process which is transparent about the composition of previous cohorts can lead to a stable matching in the long run, forming a tâtonnement process. We provide an empirical test for stability.

Using data from Australian college admissions, we first show that students have peer preferences. Our identification strategy is based on how students report preferences in a strategy-proof mechanism over college programs both across student–leveraging changes in the distribution of student abilities at programs over time–and within student–analyzing submitted preferences immediately before and after they learn how their ability relates to that of their potential peers. The patterns we observe, particularly the switching of programs in our within-person analysis, is consistent with preferences in which students do not want to be a “small fish in a big pond." They are broadly inconsistent with strategic behavior, mistakes, and changes in preferences.

We use our test for stability to show that long-lived programs in the Australian market converge to stability. Theoretically, we show that key features of the Australian market guarantee this convergence in the long run, but that entry and exit of programs causes instability for students matched to short-lived programs. This instability causes does not disipate in time, and is correlated with high attrition among students at these programs. Moreover, low socio-economic status students are particularly likely to be affected by this instability.

We propose a new mechanism to solve these issues. This mechanism is largely similar to the current matching process, except that it more closely resembles a tâtonnement process in that it does not match any students until peer preferences are (nearly) fully discovered. This mechanism is an iterative process *within each cohort*, thus removing the two types of instability discussed above. This mechanism is a relatively small modification to iterative mechanism already in use in higher-education markets in China, Brazil, Germany and Tunisia (see Bo and Hakimov (2019); Luflade (2019)).

A question that remains is: What causes peer preferences? On one hand, peer preferences could be caused by a direct aversion to being a “small fish in a big pond," or other behavioral reasons.[35] On the other hand, what we observe as peer preferences may be signs of market failures that can be addressed by market design solutions. As a simple example, suppose that seats

[35] See Dreyfuss, Heffetz, and Rabin (2019) who study how expectations-based loss aversion may lead agents not to rank otherwise desirable options in strategy-proof mechanisms. Although their model and predictions are significantly different from ours–for example, students will not “demote" programs where they are unlikely to receive admission, they will refuse to rank them outright–similar types of preferences may generate results similar to the effects we observe.

in individual classes are determined by entrance test scores. Then students may be pessimistic about the prospect of enrolling in popular classes if other students in the program have higher scores. As a result, a student may demote a prestigious program when reporting her preferences to the mechanism, even though it would be her favorite choice if she knew she would be able to enroll in certain classes, in which case a blocking pair exists. A market that resolves this uncertainty about enrolling in courses within a particular program ex ante, perhaps through a different course assignment process, may therefore improve the stability of matchings by reducing “peer preferences." Studying these microfoundations is left for future research.

References

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools.” *Econometrica* 82 (1):137–196.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda. 2015. “Expanding "Choice" in School Choice.” *American Economic Journal: Microeconomics* 7 (1):1–42.

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. “Do parents value school effectiveness?” *American Economic Review* 110 (5):1502–39.

Abdulkadiroğlu, Atila and Tayfun Sönmez. 2003. “School choice: A mechanism design approach.” *American Economic Review* 93 (3):729–747.

Allende, Claudia. 2020. “Competition Under Social Interactions and the Design of Education Policies.” *Unpublished manuscript* .

Artemov, Georgy, Yeon-Koo Che, and Yinghua He. 2020. “Strategic ‘Mistakes’: Implications for Market Design Research.” *Unpublished manuscript* .

Attewell, Paul. 2001. “The Winner-Take-All High School: Organizational Adaptations to Educational Stratification.” *Sociology of Education* 74 (4):267–295.

Azevedo, Eduardo M and Jacob D Leshno. 2016. “A supply and demand framework for two-sided matching markets.” *Journal of Political Economy* 124 (5):1235–1268.

Bagshaw, Eryk and Inga Ting. 2016. “NSW universities taking students with ATARs as low as 30.” *The Sydney Morning Herald* .

Berger, Ulrich. 2007. “Brown’s original fictitious play.” *Journal of Economic Theory* 135 (1):572–578.

Beuermann, Diether W. and C. Kirabo Jackson. 2020. “The Short and Long-Run Effects of Attending The Schools that Parents Prefer.” *Journal of Human Resources* .

Beuermann, Diether W., C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. 2019. “What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output.” *Unpublished manuscript* .

Bo, Inacio and Rustamdjan Hakimov. 2019. “The iterative deferred acceptance mechanism.” *Available at SSRN 2881880* .

Brown, George W. 1951. "Iterative Solutions of Games by Fictitious Play." In *Activity Analysis of Production and Allocation*, edited by Tjalling C. Koopmans. Wiley, 374–376.

Budish, Eric and Judd B. Kessler. 2020. "Can Market Participants Report their Preferences Accurately (Enough)?" *Unpublished manuscript* .

Bykhovskaya, Anna. 2020. "Stability in matching markets with peer effects." *Games and Economic Behavior* 122:28–54.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." *American Economic Review* 102 (6):2981–3003.

Carrasco-Novoa, Diego, Sandro Diez-Amigo, and Shino Takayama. 2021. "The Impact of Peers on Academic Performance: Theory and Evidence from a Natural Experiment." *Unpublished manuscript* .

Carroll, Gabriel. 2018. "On Mechanisms Eliciting Ordinal Preferences." *Theoretical Economics* 13 (3):1275–1318.

Chen, Li and Juan Sebastián Pereyra. 2019. "Self-selection in school choice." *Games and Economic Behavior* 117:59–81.

Chen, Yan and Tayfun Sönmez. 2002. "Improving Efficiency of On-campus Housing: An Experimental Study." *American Economic Review* 92:1669–1686.

Coles, Peter, Alexey Kushnir, and Muriel Niederle. 2013. "Preference signaling in matching markets." *American Economic Journal: Microeconomics* 5 (2):99–134.

Conley, Timothy G, Nirav Mehta, Ralph Stinebrickner, and Todd R Stinebrickner. 2018. "Social Interactions, Mechanisms, and Equilibrium: Evidence from a Model of Study Time and Academic Achievement." *CESifo Working Paper Series* .

Dobbie, Will and Roland G. Fryer Jr. 2014. "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools." *American Economic Journal: Applied Economics* 6 (3):58–75.

Dreyfuss, Bnaya, Ori Heffetz, and Matthew Rabin. 2019. "Expectations-Based Loss Aversion May Help Explain Seemingly Dominated Choices in Strategy-Proof Mechanisms." *Unpublished manuscript* .

Echenique, Federico and M. Bumin Yenmez. 2007. "A solution to matching with preferences over colleagues." *Games and Economic Behavior* 59 (1):46–71.

Ellickson, Bryan, Birgit Grodal, Suzanne Scotchmer, and William R Zame. 1999. "Clubs and the Market." *Econometrica* 67 (5):1185–1217.

Elsner, Benjamin and Ingo E. Isphording. 2017. "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." *Journal of Labor Economics* 35 (3):787–828.

Elsner, Benjamin, Ingo E. Isphording, and Ulf Zölitz. 2018. "Achievement Rank Affects Performance and Major Choices in College." *Unpublished manuscript* .

Epple, Dennis and Richard E Romano. 1998. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review* :33–62.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2019. "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions." *American Economic Review* 109 (4):1486–1529.

Frank, Robert H. 1985. *Choosing the Right Pond: Human Behavior and the Quest for Status*. Oxford University Press.

Gale, David and Lloyd S Shapley. 1962. "College admissions and the stability of marriage." *The American Mathematical Monthly* 69 (1):9–15.

Haeringer, Guillaume and Flip Klijn. 2009. "Constrained School Choice." *Journal of Economic Theory* 144 (5):1921–47.

Hassidim, Avinatan, Assaf Romm, and Ran I. Shorrer. 2020. "The Limits of Incentives in Economic Matching Procedures." *Management Science* .

James, Richard, Gabrielle Baldwin, and Craig McInnis. 1999. "Which University?: The factors influencing the choices of prospective undergraduates." *Canberra: Department of Education, Training and Youth Affairs* 99 (3).

Kojima, Fuhito, Parag A Pathak, and Alvin E Roth. 2013. "Matching with couples: Stability and incentives in large markets." *The Quarterly Journal of Economics* 128 (4):1585–1632.

Leshno, Jacob D. 2021. "Stable Matching with Peer Effects in Large Markets - Existence and Cutoff Characterization." Unpublished manuscript.

Luflade, Margaux. 2019. "The value of information in centralized school choice systems." Unpublished manuscript.

Marsh, Herbert W., Marjorie Seaton, Ulrich Trautwein, Oliver Lüdtke, K.T. Hau, Alison O'Mara, and Rhonda G. Craven. 2008. "The Big-fish–little-pond-effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research." *Educational Psychology Review* 20:319–350.

Murphy, Richard and Felix Weinhardt. 2020. "Top of the Class: The Importance of Ordinal Rank." *Review of Economic Studies, Forthcoming* .

Narita, Yusuke. 2018. "Match or mismatch? Learning and inertia in school choice." Unpublished manuscript.

Nei, Stephen and Bobak Pakzad-Hurson. 2019. "Strategic Disaggregation in Matching Markets." Unpublished manuscript.

Neilson, Christopher. 2019. "The Rise of Centralized Choice and Assignment Mechanisms in Education Markets Around the World." Unpublished manuscript.

Pycia, Marek. 2012. "Stability and Preference Alignment in Matching and Coalition Formation." *Econometrica* 80 (1):323–362.

Pycia, Marek and M. Bumin Yenmez. 2019. "Matching with Externalities." Unpublished manuscript.

Roth, Alvin E. 2002. "The economist as engineer: Game theory, experimentation, and computation as tools for design economics." *Econometrica* 70 (4):1341–1378.

Roth, Alvin E and Elliott Peranson. 1999. "The redesign of the matching market for American physicians: Some engineering aspects of economic design." *American Economic Review* 89 (4):748–780.

Rothstein, Jesse M. 2006. "Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions." *American Economic Review* 96 (4):1333–1350.

Sacerdote, Bruce. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education*, vol. 3. Elsevier, 249–277.

Scotchmer, Suzanne and Chris Shannon. 2015. "Verifiability and group formation in markets." *Available at SSRN 2662578* .

Seaton, Marjorie, Herbert W. Marsh, and Rhonda G. Craven. 2009. "Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries." *Journal of Educational Psychology* 101 (2):319–350.

Sóvágó, Sándor and Ran I. Shorrer. 2018. "Obvious Mistakes in a Strategically Simple College-Admissions Environment." *Unpublished manuscript* .

Stinebrickner, Ralph and Todd R Stinebrickner. 2006. "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds." *Journal of public Economics* 90 (8-9):1435–1454.

Wilson, Robert. 1987. "Game-Theoretic Approaches to Trading Processes." In *Advances in Economic Theory: Fifth World Congress*, edited by Truman Bewley. Cambridge University Press, 33–70.

Yu, Han. 2020. "Am I the big fish? The effect of ordinal rank on student academic performance in middle school." *Journal of Economic Behavior & Organization* 176:18–41.

Zárate, Román Andrés. 2019. "Social and Cognitive Peer Effects: Experimental Evidence from Selective High Schools in Peru."

Zhang, Luyao and Dan Levin. 2017. "Bounded Rationality and Robust Mechanism Design: An Axiomatic Approach." *American Economic Review: Papers & Proceedings* 107 (5):235–239.

Appendix

Natalie Cox Ricardo Fonseca Bobak Pakzad-Hurson

A.1 Appendix

In this section, we present proofs and figures omitted in the main text, as well as model extensions.

Theorem 1

Proof. By Lemma 2, it suffices to show the existence of a rational expectations, market clearing cutoff-distribution vector pair (p, λ). Define $Z(p, \lambda) = Z^d(p, \lambda) \times Z^\lambda(p, \lambda)$, with the first factor defined as a vector with entries for each $c \in C$ given by:

$$Z^{d,c}(p, \lambda) = \begin{cases} \frac{p^c}{1+q^c-D^c(p,\lambda)} \text{ if } D^c(p, \lambda) \leq q^c \\ p^c + D^c(p, \lambda) - q^c \text{ if } D^c(p, \lambda) > q^c \end{cases}$$

and the second defined by:

$$Z^\lambda(p, \lambda) = \lambda^x(\mu) \text{ for } \mu = A(p, \lambda)$$

Z^λ is a mapping from $[0, 1]^{N+1} \times \Lambda^{N+1}$ to Λ^{N+1} So, to summarize, function Z is a mapping from $K = [0, 1]^N \times \Lambda^{N+1} \to K$. We endow K with the product topology, and all notions of compactness and continuity will be relative to that topology.

The proof will involve the following steps:

1. If (p, λ) is a fixed point of Z, then (p, λ) is rational expectations and market clearing,
2. K is a convex, compact, non-empty Hausdorff topological vector space, and
3. Z is continuous.

The two last points imply, by Schauder fixed-point theorem, an extension of Brouwer's fixed point theorem, that Z has a fixed point, which by the first one gives our result. The formal statement of this theorem is the following:

Theorem. *(**Schauder fixed-point theorem**): Let K be a nonempty, convex, compact, Hausdorff topological vector space and let Z be a continuous mapping from K into itself. Then Z has a fixed point.*

1. To see that a fixed point (p, λ) of Z implies that (p, λ) is rational expectations and market clearing note that $Z^{\lambda}(p, \lambda) = \lambda$ implies that $\lambda = \lambda^x(\mu)$ for $\mu = A(p, \lambda)$. So (p, λ) is rational expectations. $Z^d(p, \lambda) = p$ implies that for every c either $D^c(p, \lambda) = q^c$ or $D^c(p, \lambda) \leq q^c$ and $p^c = 0$, so (p, λ) is market clearing.

2. It is clear that K is nonempty. To see convexity, we note that $[0, 1]$ is clearly convex. It remains to show that Λ is convex, which then implies the convexity of K as the product of convex sets. To see that this is the case, we must check that for two functions $\lambda, \hat{\lambda} \in \Lambda$, any function $\tilde{\lambda}$, defined as $\tilde{\lambda}(\alpha) = \beta\lambda(\alpha) + (1 - \beta)\hat{\lambda}(\alpha)$ for some $\beta \in [0, 1]$, is in Λ. To see that this is the case note that $\tilde{\lambda}(\alpha) \in [0, 1]$ for any $\alpha \in \mathcal{A}$ and any $\beta \in [0, 1]$, as $\lambda(\alpha), \hat{\lambda}(\alpha) \in [0, 1]$ and $\tilde{\lambda}(\alpha) \in [\min\{\lambda(\alpha), \hat{\lambda}(\alpha)\}, \max\{\lambda(\alpha), \hat{\lambda}(\alpha)\}]$. $\tilde{\lambda}(\cdot)$ must be also be non-decreasing; for any $x < y$ with $x, y \in [0, 1]^{N+1}$ and any $\beta \in [0, 1]$ $\tilde{\lambda}^x(\alpha) = \beta\lambda^x(\alpha) + (1 - \beta)\hat{\lambda}^x(\alpha) \leq \beta\lambda^y(\alpha) + (1 - \beta)\hat{\lambda}^y(\alpha) = \tilde{\lambda}^y(\alpha)$ where the inequality follows from the non-decreasing property of $\lambda(\cdot)$ and $\hat{\lambda}(\cdot)$.

   As $\tilde{\lambda}(\cdot)$ is a non-decreasing function from $[0, 1]$ to itself, $\tilde{\lambda} \in \Lambda$, i.e. Λ is convex.

   To show that K is compact and Hausdorff, we show that Λ is compact and Hausdorff.

   Lemma A.1. Λ *is compact and Hausdorff.*

   Proof. $[0, 1]^{[0,1]}$ is compact (in the product topology) by Tychonoff's theorem, as it is the product of compact spaces. To note the compactness of Λ it therefore suffices to show that Λ is a closed subspace of $[0, 1]^{[0,1]}$. Let $\langle\lambda(\alpha^{\ell})\rangle_{\ell=1,2,\ldots}$ be a convergent Moore-Smith sequence with limit λ, where each $\alpha^{\ell} \in \mathcal{A}$. We need to show that $\lambda \in \Lambda$. For any $x < y$ with $x, y \in [0, 1]^{N+1}$ and any ℓ it must be the case that $0 \leq \lambda^x(\alpha^{\ell}) \leq \lambda^y(\alpha^{\ell}) \leq 1$. Taking the limit with respect to ℓ yields that $0 \leq \lambda^x \leq \lambda^y \leq 1$, i.e. $\lambda \in \Lambda$. Therefore, Λ is compact.

   Similarly Λ is Hausdorff: $\Lambda \subset [0, 1]^{[0,1]}$ is Hausdorff as a subset of a Hausdorff space. □

   To complete the proof of the lemma, note that K is the product of compact, Hausdorff spaces by the previous lemma. Therefore, K is compact and Hausdorff.

3. Consider pairs (p, λ) and (p', λ') with $\mu = A(p, \lambda)$ and $\mu' = A(p', \lambda')$. Let us denote the set of students that are assigned to program c under only one of the two assignments μ, μ' by $\Delta^c(\mu, \mu') = \{\mu(c) \backslash \mu'(c)\} \cup \{\mu'(c) \backslash \mu(c)\}$. We show that the measure of $\Delta^c(\mu, \mu')$ bounds both the difference in λ and D^c between the two assignments in question:

$$\|\lambda(\mu) - \lambda(\mu')\|_{\infty} = \sup_{c,x} |\lambda^{c,x}(\mu) - \lambda^{c,x}(\mu')| \leq \max_{c} \ \eta(\Delta^c(\mu, \mu'))$$

$$\|D^c(p, \lambda) - D^c(p', \lambda')\|_{\infty} = \max_{c} |\eta(\mu(c)) - \eta(\mu'(c))| \leq \max_{c} \ \eta(\Delta^c(\mu, \mu'))$$

To see that the first inequality holds, just note that for any $c \in C$ and $x \in [0,1]^{N+1}$, $|\lambda^{c,x}(\mu) - \lambda^{c,x}(\mu')| \leq \max_c \eta(\Delta^c(\mu,\mu'))$, as for any x and c, the difference in the measure of students with scores below x at c cannot be larger than the total measure of students who are matched to c in only one of μ and μ'. For any c we have that $\eta(\Delta^c(\mu,\mu')) = \eta(\{\mu(c) \backslash \mu'(c)\}) + \eta(\{\mu'(c) \backslash \mu(c)\}) \geq |\eta(\mu(c)) - \eta(\mu'(c))|$, and therefore the second inequality holds.

Therefore, to show continuity we must show that for any $\epsilon > 0$, there exists $\delta > 0$ such that $\|p - p'\|_\infty < \delta$ and $\|\lambda(\mu) - \lambda'(\mu)\|_\infty < \delta \Rightarrow \max_c \eta(\Delta^c(\mu,\mu')) < \epsilon$. Consider a pair (p,λ) and take any $\epsilon > 0$. By Assumption A3, there is a $\delta_1 > 0$ such that if $\|\lambda(\mu) - \lambda'(\mu)\|_\infty < \delta_1$, the set of students whose preferences change, $\Delta(\lambda,\lambda') = \{\theta | \succeq^{\theta|\lambda} \neq \succeq^{\theta|\lambda'}\}$, is such that $\eta(\Delta(\lambda,\lambda')) < \epsilon/2$. Take now $\Delta(p,p')$, the set of students who can be admitted to some program c under one of p, p' but not under the other, so that $\Delta(p,p') = \cup_c \{\theta | \min\{p^c, p'^c\} \leq r^{\theta,c} < \max\{p^c, p'^c\}\}$. Then

$$\Delta^c(\mu,\mu') \subset \Delta(p,p') \cup \Delta(\lambda,\lambda') \tag{A.1}$$

For $\delta_2 < \epsilon/2N$, we have that if $\|p - p'\| < \delta_2$, then $\eta(\Delta(p,p')) \leq \sum_c |p^c - p'^c| \leq N \cdot \epsilon/2N = \epsilon/2$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then if (p,λ) and (p',λ') satisfy $\|p - p'\|_\infty < \delta$ and $\|\lambda(\mu) - \lambda'(\mu)\|_\infty < \delta$, we have that

$$\eta(\Delta(p,p') \cup \Delta(\lambda,\lambda')) \leq \eta(\Delta(p,p')) + \eta(\Delta(\lambda,\lambda')) < \epsilon$$

By Equation A.1, this implies $\eta(\Delta(\mu,\mu')) < \epsilon$. Thus Z is continuous and our proof concludes.

□

Proposition 1

Proof.

1. Let μ_* be a stable matching. For each θ, let $\tilde{\succ}^\theta$ be such that $\mu_*(\theta)$ is the unique acceptable program. Because φ is stable, $\varphi(\tilde{\succ}) = \mu_*$. To see that this is a Nash equilibrium, note that for any θ, and any program $c \succ^{\theta|\mu_*} \mu_*(\theta)$, stability of μ_* implies that there is no report $\succ^\theta$ that will result in θ matching with c.

   Suppose for contradiction that $\tilde{\succ}$ is a Nash equilibrium of φ but that $\mu = \varphi(\tilde{\succ})$ is not a stable matching. Then there exists some $\theta \in \Theta$ and some $c \in C$ such that (θ, c) form a blocking pair (with respect to $\succeq^{\theta|\mu}$). By Remark 2 and the fact that φ is a stable mechanism, μ is the unique stable matching with respect to the submitted preferences $\tilde{\succ}$. Let p be the associated cutoff vector. Now consider reported preferences $\hat{\succ}$ where $\hat{\succ}^{\theta'} = \tilde{\succ}^{\theta'}$ for all $\theta' \neq \theta$ and $\hat{\succ}^{\theta}$ lists only program c as acceptable. There is similarly a unique stable matching μ' with respect to these preferences, but the cutoff vector for this stable matching must also be p, due to the reported preferences of a zero measure set of students differing between $\hat{\succ}$ and $\tilde{\succ}$. Since (θ, c) block μ it must be that $r^{\theta,c} \geq p^c$. But then $\varphi^{\theta}(\hat{\succ}) = c$ since c is a stable mechanism. Contradiction with $\tilde{\succ}$ being a Nash equilibrium.

2. Let $\tilde{\succ}$ be a Bayes Nash equilibrium, and suppose for contradiction that $\varphi(\tilde{\succ}) = \mu_*$. By Remark 2 and the ongoing assumption that μ_* is stable, it must be that μ_* is associated with some cutoff vector p, and by assumption **A2** it must be that $p^c < 1 - q^c$ for all $c \in C \setminus \{c_0\}$.

   Consider the set of students with scores $r^{\theta} > \mathbf{1} - q$ who lack rationality for the top choice at $\tilde{\succ}$, $L_{\tilde{\succ},\mathbf{1}-q}$. Recall that we have assumed $\eta(L_{\tilde{\succ},\mathbf{1}-q}) > 0$. Because φ respects rankings, it must be that any student type $\theta \in L_{\tilde{\succ},\mathbf{1}-q}$ believes with probability one that $\varphi^{\theta}(\tilde{\succ}) = \mu_*(\theta)$ is the $\tilde{\succ}^{\theta}$-maximal program. By the equilibrium hypothesis, it must be that the $\mu_*(\theta)$ is the $\succeq^{\theta|\sigma,\tilde{\succeq}}$-maximal program.

   By the stability of μ_* and the fact that $r^{\theta} > \mathbf{1} - q$ for all $\theta \in L_{\tilde{\succeq},\mathbf{1}-q}$ it must also be the case that $\mu_*(\theta)$ is the $\succeq^{\theta|\mu_*}$-maximal program for all $\theta \in L_{\tilde{\succ},\mathbf{1}-q}$. Therefore, our arguments imply that the top-ranked program according to $\succeq^{\theta|\sigma,\tilde{\succ}}$ is the same as the top-ranked program according to $\succeq^{\theta|\mu_*}$ for all $\theta \in L_{\tilde{\succ},\mathbf{1}-q}$. But by the assumption that any $\theta \in L_{\tilde{\succ},\mathbf{1}-q}$ lacks rationality for the top choice at $\tilde{\succ}$, the two respective top-ranked programs must differ. Contradiction.

□

Theorem 2

Proof.

1. **"If" part** If μ_* is stable we know that it satisfies rational expectations, so $S(p_*, \lambda_*) = \lambda(A(p_*, \lambda_*)) = \lambda_*$, and therefore λ_* (and also (p_*, λ_*)) is in steady-state.

   "Only if" part Take an ability distribution in steady-state λ_*. Then $\lambda_* = S(P\lambda_*, \lambda_*)$, so $(P\lambda_*, \lambda_*)$ satisfy rational expectations. By the definition of P, $P\lambda_*$ is market clearing given λ_*. Therefore, by Lemma 2, we know that $\mu_* = A(P\lambda_*, \lambda_*)$ is stable.

2. **"If" part** Take any $\epsilon > 0$ and λ_{t-1}. Given that $\mu_t \in M$, θ is involved in at least one blocking pair at μ_t if and only if $D^{\theta}(p_t, \lambda_t) \neq D^{\theta}(p_t, \lambda_{t-1})$; if (θ, c) block μ_t then $r^{\theta,c} \geq p^c$ and $c \succeq^{\theta|\mu_t} \mu_t(\theta)$, implying that $D^{\theta}(p_t, \lambda_t) \neq D^{\theta}(p_t, \lambda_{t-1})$, and if $D^{\theta}(p_t, \lambda_t) \neq D^{\theta}(p_t, \lambda_{t-1})$ then $(\theta, D^{\theta}(p_t, \lambda_t))$ block μ_t. Some subset of students whose ordinal rankings change can block μ_t; $\eta(\{\theta | (\theta, c) \text{ block } \mu_t \text{ for some } c \in C\}) \leq \eta(\{\theta | \succeq^{\theta|\mu_{t-1}} \neq \succeq^{\theta|\mu_t}\})$. By **A4**, there exists $\delta > 0$ such that $\|\lambda_{t-1} - \lambda_t\|_{\infty} < \delta$ implies $\eta(\{\theta | \succeq^{\theta|\mu_{t-1}} \neq \succeq^{\theta|\mu_t}\}) < \epsilon$. Therefore, for $\|\lambda_{t-1} - \lambda_t\|_{\infty} < \delta$, $\eta(\{\theta | (\theta, c) \text{ block } \mu_t \text{ for some } c \in C\}) \leq \epsilon$ as desired.

"Only if" part Fix $\delta > 0$ and λ_{t-1}, and let B be the set of student types involved in at least one blocking pair at μ_t. Take three economies $E_t = [\zeta^{\eta,\mu_{t-1}}, q], E_{t+1} = [\zeta^{\eta,\mu_t}, q]$, and $E' = [\zeta', q]$, where measure ζ' is defined as follows: for any open set $R \subset [0,1]^{N+1}$, any matching ν, and any $\succeq \in P$, $\zeta'(\{\theta | r^\theta \in R \cap B \text{ and } \succeq^{\theta|\nu} = \succeq\}) = \eta(\{\theta | r^\theta \in R \cap B \text{ and } \succeq^{\theta|\mu_{t-1}} = \succeq\})$ and $\zeta'(\{\theta | r^\theta \in R \cap \{\Theta \setminus B\} \text{ and } \succeq^{\theta|\nu} = \succeq\}) = \eta(\{\theta | r^\theta \in R \cap \{\Theta \setminus B\} \text{ and } \succeq^{\theta|\mu_t} = \succeq\})$. That is, ζ' specifies student types such that students involved in blocking pairs have the same preferences as in economy E_t and students not involved in blocking pairs have the same preferences as in economy E_{t+1}. Let μ_t, μ_{t+1}, and μ' be the stable matchings in each of these economies, respectively. Recall that by Remark 2, μ_t and μ_{t+1} are the outcomes of the TIM procedure at times t and $t+1$, respectively.

We claim that $\mu_t = \mu'$. To see this, note that $\theta \in B$ if and only if $D^\theta(p_t, \lambda_t) \neq D^\theta(p_t, \lambda_{t-1})$.[1] Then as $\epsilon \to 0$, $\zeta' \to \zeta^{\eta,\mu_t}$ in the weak-* sense. By the earlier argument that $\mu_t = \mu'$ and Lemma B3 of Azevedo and Leshno (2016), this implies that $\eta(\{\theta | \mu_t(\theta) \neq \mu_{t+1}(\theta)\}) \to 0$. Therefore, $\epsilon \to 0$ implies $\|\lambda_{t-1} - \lambda_t\|_\infty \to 0$.

□

Local Convergence

We show that the TIM procedure does not necessarily exhibit local convergence.

Definition 7. *A stable matching $\mu_* = (p_*, \lambda_*)$ is* locally convergent *if for any $\epsilon > 0$ there exists $\delta > 0$ and $T > 0$ such that for any λ_0 satisfying $||\lambda_0 - \lambda_*||_\infty < \delta$ and any $t > T$, $||\mu_* - \mu_t||_\infty < \epsilon$.*

This is a weaker notion of convergence, since we restrict ourselves to starting distributions λ_0 that are "close to" the stable matching distribution. Practically, if a stable matching satisfies this condition, then we are guaranteed to create a stable matching in the long run if the initial beliefs in the student distribution at each program is close to that in a stable matching.

Remark 7. *For a stable matching μ_* in some economy E, local convergence is not guaranteed, even if μ_* is the unique stable matching in economy E.*

Proof. We prove this remark via the following example.

Example 4. *There is one program c with $q \geq 1$, and $r^{\theta,c} = r^{\theta,c_0} = r^\theta$ for all $\theta \in \Theta$. Let $s(\lambda(\mu))$ be the mean score r^θ of students assigned to c in μ, that is*

$$s(\lambda) = \frac{1}{\lambda^{c,(1,1)}} \int_0^1 y d\lambda^{c,(y,y)}$$

[1] If $\theta \in B$ then let $c \in C$ be θ's most preferred program (according to $\succeq^{\theta|\mu_t}$) with which she blocks μ_t. Then $c = D^\theta(p_t, \lambda_t) \neq D^\theta(p_t, \lambda_{t-1})$. If $D^\theta(p_t, \lambda_t) \neq D^\theta(p_t, \lambda_{t-1})$ then $(\theta, D^\theta(p_t, \lambda_t))$ is a blocking pair.

Each student θ receives zero utils from remaining unmatched. $\gamma < 1$ fraction of students have weak peer preferences and receive strictly positive utility from attending c regardless of λ. Students with weak peer preferences have scores r^θ that are uniformly distributed. The remaining $1-\gamma$ have strong peer preferences and receive utility $v^\theta - f(s(\lambda), r^\theta)$ from matching with the program, where

$$f(s(\lambda), r^\theta) = \begin{cases} 0 \text{ if } r^\theta \geq \frac{1}{2} \text{ and } s(\lambda) \leq \frac{1}{2} \\ 0 \text{ if } r^\theta < \frac{1}{2} \text{ and } s(\lambda) > \frac{1}{2} \\ k|\frac{1}{2} - s(\lambda)| \text{ otherwise} \end{cases}$$

A student θ is better off enrolling at the program if and only if $v^\theta - f(s(\mu), r^\theta) \geq 0$, where we break ties in favor of the student attending the program. Let each v^θ and each r^θ be distributed independently and uniformly over $[0,1]$. The peer preference term $f(\cdot,\cdot)$ reflects that students want their own score to be different from the average scores of their peers, and suffer loss proportional to the average score of students if they are in the "majority" type.

Let $\mu_*^\theta = c$ for all $\theta \in \Theta$, which is a matching since $q^c \geq 1$. Then $\lambda_* = \lambda(\mu^*)$ has the property that $\lambda_*^{(y,y)} = y$ for all $y \in [0,1]$. We first note that $\mu_* = A(0, \lambda^*)$ is stable: it is market clearing (i.e. $p_* = 0$) and satisfies rational expectations, i.e. $s(\lambda_*) = \frac{1}{2}$ and so all students attend c. Furthermore, it is easy to see that this is the unique stable matching. Any market clearing matching $\mu_{\prime}$ must satisfy $p_{\prime} = 0$. If $s_{\prime} = s(\lambda(\mu_{\prime})) < \frac{1}{2}$ all the students with scores $r^\theta > \frac{1}{2}$ prefer to be matched to c while only a fraction of the students with scores $r^\theta \leq \frac{1}{2}$ prefer to be matched to c. This implies that $s(\lambda(A(p_{\prime}, s_{\prime}))) > \frac{1}{2} > s_{\prime}$. Therefore, $(p_{\prime}, \lambda(\mu_{\prime}))$ does not satisfy rational expectations, and so $\mu_{\prime}$ is not stable. A similar argument follows if $s_{\prime} > \frac{1}{2}$.

We claim that the TIM procedure does not converge for any $s_0 = s(\lambda(\mu_0)) \neq \frac{1}{2}$ when $k \geq \frac{8}{1-\gamma}$. Recall that as $s(\cdot)$ is a function of λ, if the sequence $s_1, s_2, ...$ does not converge, then neither does the sequence $\lambda_1, \lambda_2, ...$

To show this claim, let $s_0 = \frac{1}{2} - \delta$ for some $\delta > 0$ (by the symmetry of the market, similar logic holds if $\delta < 0$). First suppose that $k\delta \geq 1$. Then in μ_1, none of the students with $r^\theta < \frac{1}{2}$ who have strong peer preferences will enroll in c, and all other student types will. Therefore,

$$s(\lambda(\mu_1)) = \frac{\frac{1}{4}(\frac{1}{2}\gamma) + \frac{3}{4}\frac{1}{2}}{\frac{1}{2}(1+\gamma)} = \frac{3+\gamma}{4(1+\gamma)}$$

Similarly,

$$s(\lambda(\mu_2)) = \frac{1+3\gamma}{4(1+\gamma)}$$

From there, a cycle forms: for any odd $t > 1$, $s(\lambda(\mu_t)) = s(\lambda(\mu_1))$ and $s(\lambda(\mu_{t+1})) = s(\lambda(\mu_2))$, meaning that the market does not converge to the unique stable matching.

Now suppose $k\delta < 1$. By a similar calculation, we have that

$$s(\lambda(\mu_1)) = \frac{\gamma + (1-\gamma)(1-k\delta) + 3}{4(1+\gamma+(1-\gamma)(1-k\delta))}$$

For $k \geq \frac{8}{1-\gamma}$ we claim that $s(\lambda(\mu_1)) \geq \frac{1}{2} + \delta$. To see this, note that $\frac{\gamma+(1-\gamma)(1-k\delta)+3}{4(1+\gamma+(1-\gamma)(1-k\delta))} - \frac{1}{2} - \delta \geq 0$ if and only if $k\delta - \gamma k\delta - 8\delta + 4k\delta^2 - 4\gamma k\delta^2 \geq 0$. Since $\gamma < 1$, $k\delta - \gamma k\delta - 8\delta \geq 0$ implies the desired condition.

Noting the symmetry of the market, it is the case that for odd t, the sequence s_t, s_{t+2}, s_{t+4}... is non-decreasing where each element is strictly larger than $\frac{1}{2}$ and $s_{t+1}, s_{t+3}, s_{t+5}, \ldots$ is non-increasing where each element is strictly smaller than $\frac{1}{2}$. Therefore, the TIM process does not converge.

□

Proposition 4

Proof.

1. If the TIM procedure converges to $\mu_* = A(p_*, \lambda_*)$, then for any stopping rule $\delta > 0$ the TFM mechanism must terminate, and we show that we can pick $\delta > 0$ such that at the stopping step of the TFM mechanism $\tau(\delta)$, $\lambda_{\tau(\delta)-1}$ is arbitrarily close to λ_*. To see this, fix any $\gamma > 0$. In the TIM procedure, there exists $\tau(\gamma) \geq 0$ such that $||\lambda_* - \lambda_\tau||_\infty < \gamma$ for all $\tau \geq \tau(\gamma)$ by the assumption that the TIM procedure converges to μ_*. Let $\Delta_\tau := ||\lambda_\tau - \lambda_{\tau-1}||_\infty$, $\tau > 0$. It must be that $\Delta_\tau > 0$ for all τ such that $\lambda_* \neq \lambda_\tau$. Moreover, $\Delta_\tau \to 0$ i.e. the sequence $\lambda_1, \lambda_2, \ldots$ must be Cauchy because it is convergent. Let $\delta \in (0, \min\limits_{\tau \leq \tau(\gamma)} \Delta_\tau)$. Then the TFM mechanism must terminate at some $\tau \geq \tau(\gamma)$.

   For stopping time $\tau(\delta)$ the final matching in the TFM mechanism is $\mu_{\mu_0,\delta}(\theta) = A(P\lambda_{\tau(\delta)-1}, \lambda_{\tau(\delta)-1})$. Therefore, the previous argument completes our proof if we can show that for any $\epsilon > 0$ there exists $\gamma > 0$ such that if $||\lambda_* - \lambda_{\tau(\delta)-1}||_\infty < \gamma$, then $\eta(\{\theta | \mu_{\mu_0,\delta}(\theta) = \mu_*(\theta)\}) > 1 - \epsilon$.

   To show this, first note that for any $\epsilon_1 > 0$ there exists some sufficiently small $\gamma_1 > 0$ such that if $\|\lambda_* - \lambda_{\tau(\delta)-1}\|_\infty < \gamma_1$ then $\eta(\{\theta | \succeq^{\theta|\lambda_*} \neq \succ^{\theta|\lambda_{\tau(\delta)-1}}\}) < \epsilon_1$, by Assumption A4.

   Second, we show that for any $\epsilon_2 > 0$ there exists $\gamma_2 > 0$ such that $\|p_* - p_{\tau(\delta)}\|_\infty < \epsilon_2$ if $\|\lambda_* - \lambda_{\tau(\delta)-1}\|_\infty < \gamma_2$. Consider two economies $E_{\tau(\delta)-1}$ and E_* where program rankings over students are identical, and student preferences over programs absorb the utility effects of peers characterized by λ_* and $\lambda_{\tau(\delta)-1}$, respectively. That is, $E_{\tau(\delta)-1} = [\zeta^{\eta,\mu_{\tau(\delta)-1}}, q]$ and $E_* = [\zeta^{\eta,\mu_*}, q]$. By Remark 2, there exists a unique stable matching in each of $E_{\tau(\delta)-1}$ and E_*, and these are $\mu_{\tau(\delta)}$ and μ_*, respectively. Theorem 2 of Azevedo and Leshno (2016) implies continuity of the unique stable matching of a non-peer preferences economy in student preferences. That is, for μ_* and any $\epsilon_2 > 0$ there exists some sufficiently small $\gamma_2 > 0$

such that the market clearing cutoffs in the two economies satisfy $\|p_* - p_{\tau(\delta)}\|_\infty < \epsilon_2$ when $\|\lambda_* - \lambda_{\tau(\delta)-1}\|_\infty < \gamma_2$.

A student type θ is matched to a different program in the stable matchings for the two different markets, $\mu_*(\theta) \neq \mu_{\tau(\delta)}(\theta)$, only if one of the following conditions hold: either her preferences differ ($\succ^{\theta|\mu_{\tau(\delta)-1}} \neq \succeq^{\theta|\mu_*}$), or the set of programs to which she can gain entry differ (there exists c such that $p^c_{\tau(\delta)} \leq r^{\theta,c} < p^c_*$ or $p^c_{\tau(\delta)} > r^{\theta,c} \geq p^c_*$). For any ϵ, let $\epsilon_1 + (N+1)\epsilon_2 < \epsilon$. We have shown that for $\gamma = \min\{\gamma_1, \gamma_2\}$ the former set of students has measure strictly smaller than ϵ_1, and for each c the measure of students in the latter set is strictly smaller than ϵ_2, and because there are $N+1$ programs, the measure of the latter set is strictly smaller than $(N+1)\epsilon_2$. Since $\epsilon > \epsilon_1 + (N+1)\epsilon_2$, we arrive at the desired result for $\gamma = \min\{\gamma_1, \gamma_2\}$.

2. Fix $\epsilon > 0$ and a stable matching μ_*. The proof of point 1 of the current result implies that there exists $\gamma_1 > 0$ such that $\|\lambda_1 - \lambda_0\|_\infty < \delta$ when $\|\lambda_* - \lambda_0\|_\infty < \gamma_1$. Therefore, $\mu_{(\mu_0,\delta)} = A(P\lambda_0, \lambda_0)$ for any μ_0 with $\|\lambda_* - \lambda_0\|_\infty < \gamma_1$. For any such μ_0, the proof of point 1 of the current result additionally implies that there exists $\gamma_2 > 0$ such that if $||\lambda_* - \lambda_0||_\infty < \gamma_2$, then $\eta(\{\theta|\mu_{\mu_0,\delta}(\theta) = \mu_*(\theta)\}) > 1 - \epsilon$. Therefore, the outcome of the TFM mechanism is $\epsilon-$stable for any stopping criterion δ if $||\lambda_* - \lambda_0||_\infty < \min\{\gamma_1, \gamma_2\}$.

   Example 4 shows constructively an example of a market such that $\eta(\{\theta|\mu_{\mu_0,\delta}(\theta) = \mu_*(\theta)\}) > 1 - \epsilon$ for any λ_0 sufficiently close to λ_*, whereas the TIM procedure will not converge for any $\lambda_0 \neq \lambda_*$.

3. Suppose the TFM mechanism terminates in period τ. Because the final matching is not constructed in any steps steps $k < \tau$ when λ_k is being updated, and because each λ_k is unaffected by the submitted preferences of any zero measure set of student, no student affects the final matching by lying in any step $k < \tau$. Therefore, we only regard the case in which the TFM mechanism terminates for (μ_0, δ), and consider incentives to misreport at the final step.

   Fix $\epsilon > 0$. Termination implies that $||\lambda_\tau - \lambda_{\tau-1}||_\infty < \delta$. By Assumption **A4**, for sufficiently small δ this implies that $\eta(\{\theta| \succeq^{\theta|\mu_\tau} \neq \succ^{\theta|\mu_{\tau-1}}\}) < \epsilon$. Assuming (almost) all students $\theta' \in \Theta \setminus \{\theta\}$ report preferences $\succeq^{\theta'|\mu_{\tau-1}}$, we have that θ can profitably misreport her preferences only if $\succeq^{\theta|\mu_\tau} \neq \succ^{\theta|\mu_{\tau-1}}$. Therefore, $\eta(\Theta') < \epsilon$.

4. Suppose that the TFM mechanism terminates at step $\tau = K \cdot T + t$. Note that the stopping criterion is independent of K, T, t. Therefore, we can treat τ as a constant. For sufficiently large T, $K = 0$ and $\tau = t$. Moreover, $\frac{t}{T} = \frac{\tau}{T}$ is arbitrarily small. $K = 0$ implies that no student reports her preferences more than twice, and $t = \tau$ implies that the share of submarkets who report preferences twice is $\frac{\tau}{T}$. Recall our assumption that $\eta(\Theta_k) \to 0$ for all $k \in 1, ..., T$ as

$T \to \infty$. Therefore, for any ϵ there exists T such that

$$\sum_{k=1}^{\tau} \eta(\Theta_k) < \epsilon.$$

□

Proposition 2

Proof. Fix a market E_t. We first construct a stable matching and then prove that it is unique. The algorithm for finding it proceeds in a series of steps $\ell = 1, 2, 3, \ldots$. It begins with all students facing zero peer costs from all programs, and selecting their favorite programs. As the algorithm progresses, the summary statistics for programs become "locked in" and students internalize the associated peer costs in subsequent steps.

Step 1: Begin with the matching μ_0 wherein $\mu_0(\theta) = c_0$ for all $\theta \in \Theta$. Therefore, $s_0 = s(\lambda(\mu_0))$ is the zero vector. Let $\nu_1 = A_t(P_t\lambda(\mu_0), \lambda(\mu_0))$ be the unique market clearing matching corresponding to s_0. Let $C_t^1 = \{c \in C_t | s^c(\lambda(\nu_1)) \geq s^{c'}(\lambda(\nu_1)) \forall c' \in C_t\}$. Let $D_t^1 = C_t \setminus C_t^1$. Construct matching μ_1, where $\mu_1(\theta) = \nu_1(\theta)$ if $\nu_1(\theta) \in C_t^1$ and $\mu_1(\theta) = c_0$ otherwise. Therefore, $s_1^c = s^c(\lambda(\nu_1))$ for all $c \in C_t^1$ and $s_1^{c'} = 0$ for all $c' \in D_t^1$.

Step ℓ: Begin with $s_{\ell-1}$ as defined in Step $\ell - 1$ and let $\nu_\ell = A_t(P_t\lambda(\mu_{\ell-1}), \lambda(\mu_{\ell-1}))$ be the unique market clearing matching corresponding to $s_{\ell-1}$. Let $C_t^\ell = \{c \in D_t^{\ell-1} | s^c(\lambda(\nu_\ell)) \geq s^{c'}(\lambda(\nu_\ell)) \forall c' \in D_t^{\ell-1}\}$. Let $D_t^\ell = D_t^{\ell-1} \setminus C_t^\ell$. Construct matching μ_ℓ, where $\mu_\ell(\theta) = \nu_\ell(\theta)$ if $\nu_1(\theta) \in C_t \setminus D_t^\ell$, and $\mu_\ell(\theta) = 0$ otherwise. Therefore, $s_\ell^c = s^c(\lambda(\nu_\ell))$ for all $c \in C_t \setminus D_t^\ell$ and $s_\ell^{c'} = 0$ for all $c' \in D_t^\ell$

Terminate after the (first) step ℓ' in which $D_t^{\ell'}$ is empty and let $\mu_t^{SD} = \mu_\ell$.

Note that the algorithm must terminate in at most $N + 1$ steps, as at each step ℓ at least one program is removed from D_t^ℓ.

We first show (by induction) the following result on the above algorithm:

Lemma A.2. *If $c \in C_t^\ell$ for some ℓ then $s_\ell^c = s_{\ell^*}^c$ for all $\ell^* > \ell$.*

Proof.

Base case: Show $s_1^c = s_2^c$ for all $c \in C_t^1$.

No student θ with $r^\theta \geq s_1^c$ faces peer costs from any program $c \in C_t^1$ in matching μ_1. Therefore, all such students will attend the same program in steps 1 and 2, i.e. $\mu_1(\theta) = \mu_2(\theta)$ for all $\theta \in \Theta$ with $r^\theta \geq s_1^c$. As there is a k^c measure of students matched to program c with scores higher than s_1^c, $\eta(\{\theta \in \mu_1(c) | r^\theta \geq s_1^c\}) = \eta(\{\theta \in \mu_2(c) | r^\theta \geq s_1^c\}) = k^c$ (or 0 if $\eta(\{\theta \in \mu_1(c)\}) = \eta(\{\theta \in \mu_2(c)\}) < k^c$). Therefore, $s_1^c = s_2^c$.

Induction step: Assume $s^c_{\ell-1} = s^c_\ell$ for all $c \in C_t \setminus D^{\ell-1}_t$. Show $s^c_\ell = s^c_{\ell+1}$ for all $c \in C_t \setminus D^\ell_t$.

> No student θ with $r^\theta \geq s^c_\ell$ faces peer costs from any program $c \in C^\ell_t$ in matching μ_ℓ. Moreover, by the induction hypothesis, every student θ with $r^\theta \geq s^c_{\ell-1}$ faces the same peer costs from any program $c \in C_t \setminus D^{\ell-1}_t$. Therefore, each student θ with $r^\theta \geq s^c_\ell$ will attend the same program in steps ℓ and $\ell+1$, i.e. $\mu_\ell(\theta) = \mu_{\ell+1}(\theta)$ for such students. As there is a k^c measure of students matched to each program $c \in C_t \setminus D^\ell_t$ with scores higher than s^c_ℓ, $\eta(\{\theta \in \mu_\ell(c) | r^\theta > s^c_\ell\}) = \eta(\{\theta \in \mu_{\ell+1}(c) | r^\theta > s^c_{\ell+1}\}) = k^c$ (or 0 if $\eta(\{\theta \in \mu_\ell(c)\}) = \eta(\{\theta \in \mu_{\ell+1}(c)\}) < k^c$). Therefore, $s^c_\ell = s^c_{\ell+1}$.

□

We return to the proof of the proposition.

Proof of stability of μ^{SD}_t If the terminating step of the algorithm is ℓ', then by construction $\mu^{SD}_t = \mu_{\ell'} = \nu_{\ell'}$ since $D^{\ell'}$ is empty. Therefore, $\mu^{SD}_t = A_t(P_t\lambda(\mu_{\ell'-1}), \lambda(\mu_{\ell'-1}))$, and so μ^{SD}_t is market clearing. Moreover, because $D^{\ell'}$ is empty it is the case that had we run the algorithm for one more step, $\mu_{\ell'} = \mu_{\ell'+1}$ by our induction argument, implying $\mu_{\ell'+1} = A_t(P_t\lambda(\mu_{\ell'}), \lambda(\mu_{\ell'}))$. Therefore, $\lambda(\mu_{\ell'}) = \lambda(A_t(P_t\lambda(\mu_{\ell'}), \lambda(\mu_{\ell'})))$, and so $\mu^{SD}_t = \mu_{\ell'}$ is also market clearing. By Lemma 2, μ^{SD}_t is stable.

Proof of uniqueness

To show that μ^{SD}_t is the unique stable matching, it suffices to show that $s_{SD} = s(\lambda(\mu^{SD}_t))$ is the unique stable-matching summary statistic vector. Suppose for contradiction that there exists a distinct stable-matching summary statistic vector s_*. Let K represent the subset of programs that have different summary statistics in the two stable matchings, that is, $K = \{c | s^c_{SD} \neq s^c_*\}$. By the assumption of the existence of s_* we know that K is non-empty. WLOG suppose that $K = \{c_1, c_2, ..., c_{|K|}\}$. Let $s^{max} = \max\{s^{c_1}_{SD}, s^{c_1}_*, s^{c_2}_{SD}, s^{c_2}_*, ..., s^{c_{|K|}}_{SD}, s^{c_{|K|}}_*\}$, and let $c^{max} \in \{c | s^c_{SD} = s^{max} \text{ or } s^c_* = s^{max}\}$. In words, s^{max} is the largest summary statistic that differs between the two matchings, and c_{max} is (one of) the program that has this summary statistic in one of the two matchings.

Consider the set of students $I^{max} = \{\theta | r^\theta \geq s^{max}\}$. Note that the mass of students within I^{max} enrolled at c_{max} is strictly lower than $k^{c_{max}}$ in exactly one of μ_{SD} and μ_* and is exactly equal to $k^{c_{max}}$ in the other. We claim that almost all $\theta \in I^{max}$ must be matched to the same program in both matchings, $\mu_{SD}(\theta) = \mu_*(\theta)$ for almost all $\theta \in I^{max}$. This claim will complete the contradiction. To see this, note that $f^\theta(r^\theta, s^c_{SD}) = f^\theta(r^\theta, s^c_*)$ for all $\theta \in I^{max}$: each such student θ faces the same peer cost from programs with higher summary statistics than s^{max} because these summary statistics are identical in both matchings by the definition of s^{max}, and θ faces 0 peer costs from all other programs c_i, as $r^\theta > s^{c_i}_{SD}$ and $r^\theta > s^{c_i}_*$. By Assumption **A1**, only a zero measure set of students in I^{max} could receive different matchings without forming blocking pairs. But if almost all $\theta \in I^{max}$ receive the same matching, this contradicts the ongoing assumption that program c_{max} fills exactly

k^c measure of seats from students $\theta \in I^{max}$ in one of the “stable" matchings, but it fills strictly fewer measure seats in the other “stable" matching. Therefore, there cannot exist distinct stable-matching summary statistic vectors.

Proof of Bullets 1.-3.

1. It suffices to show that there is no step ℓ in the above algorithm such that $c \in C_t \cup B_2$ is an element of C_t^ℓ and $c' \in B_1$ is an element of D_t^ℓ. If this were the case, then by Lemma A.2, $s_{SD}^c > s_{SD}^{c'}$. But then by **AA2**, all students face weakly larger peer costs from c than from c'. By **AA5**, all students must therefore prefer c' to c at matching μ_{SD}. But that contradicts that $s_{SD}^c > s_{SD}^{c'}$.

2. This follows from the first bullet, and a nearly-identical argument to the proof of uniqueness.

3. This follows from the first two bullets and assumptions **AA2** and **AA5**.

□

Theorem 3

Proof. We show that in the TIM process, the summary statistics of all programs (in B_1) exactly reach those in the stable matching in finite time. The following roapmap outlines our proof approach.

Scenario 1 : All programs are part of the first block, so that $B_1 = C \setminus \{c_0\}$

We first consider an economy with no entry and exit (**Scenario 1**). The proof is done by induction on the index of programs, ordered by their stable statistics in the unique stable matching (given no entry and exit, there is only one stable matching over time).

The **Base Case** states that there will be a period in which c_1's summary statistic reaches its stable matching value $s_*^{c_1}$, and that it will remain at this level for all future periods. This is done by (**Claim 1**) showing that if $s_t^{c_1}$ ever falls weakly below $s_*^{c_1}$, then $s_{t'}^{c_1} = s_*^{c_1}$ for all $t' > t$. **Claim 2** shows that the maximum summary statistic among all programs cannot always lie above $s_*^{c_1}$. This completes the proof of the **Base Case**. The argument for the **Induction Step** is similar, noting that all programs converge to their stable summary statistics "from the top."

Scenario 2 : Not all programs are in block B_1**, so that** $B_1 \subsetneq C \setminus \{c_0\}$ (**Scenario 2**) extends the result to cases with entry and exit. We show that the convergence for statistics of programs in B_1 occurs regardless of the entry and exit of programs in B_2, as these programs are always less preferred to ones in B_1 for high-scoring students.

We now being the proof of the first scenario.

Scenario 1: All programs are part of the first block, so that $B_1 = C \setminus \{c_0\}$

Let $\mu^* = A(p_*, s_*)$ be the unique stable matching, and suppose WLOG that $s_*^{c_1} \geq s_*^{c_2} \geq ... \geq s_*^{c_N}$. Note that as there is no exit and entry in the current scenario (as all programs are in B_1), we can omit time indexes t for this unique stable matching. We consider the generic case in which some (possibly) empty subset of programs $C' \subset C$ satisfy $s_*^{c_j} = 0$ if and only if $c_j \in C'$ and $s^{c_i} > s^{c_{i+1}}$ for $c_i \notin C'$. The proof is by induction on the index of the programs.

We first address programs $c_i \notin C'$, i.e. those for which $s_*^{c_i} > 0$.
Base Case: If $c_1 \notin C'$, there exists t such that for all $t' > t$, $s_{t'}^{c_1} = s_*^{c_1}$ and $s_*^{c_1} > s_t^{c_i}$ for all $c_i \neq c_1$.

Proof.
Claim 1: If $s_t^{c_i} \leq s_*^{c_1}$ for every program $c_i \in C$, then $s_{t+1}^{c_1} = s_*^{c_1}$ and $s_*^{c_1} > s_{t+1}^{c_i}$ for all $c_i \in C$.

Proof. As $s_t^{c_i} \leq s_*^{c_1}$ for every program c_i, (almost) all student types θ with $r^\theta \geq s_*^{c_1}$ will satisfy $\mu_{t+1}(\theta) = \mu_*(\theta)$. This is because $\succeq^{\theta|s_t} = \succeq^{\theta|s_*}$ for all such θ because they face no peer costs at any programs given s_t and s_*.

c_1 will therefore enroll exactly k^{c_1} measure of students with scores $r^\theta \geq s_*^{c_1}$, and each $c_j \neq c_1$ will enroll strictly fewer than k^{c_j} measure of students with scores $r^\theta \geq s_*^{c_1}$ by virtue of the fact that $s_*^{c_1} > s_*^{c_j}$. Therefore, $s_{t+1}^{c_1} = s_*^{c_1}$ and $s_*^{c_1} > s_{t+1}^{c_j}$ for all $c_j \neq C_1$. □

Returning to the proof of the base case, from Claim 1 we know that if $s_t^{c_i} \leq s_*^{c_1}$ for all c_i and some t then we are done. To show this, we assume for contradiction that there is no t such that $s_t^{c_i} \leq s_*^{c_1}$ for all c_i. Let $s_t^m = \max_{c_i} s_t^{c_i}$. Therefore, the condition that there is no t such that $s_t^{c_i} \leq s_*^{c_1}$ for all c_i is equivalent to $s_t^m > s_*^{c_1}$ for all t.
Claim 2: If $s_t^m > s_*^{c_1}$ for all t, then $s_1^m > s_2^m > ....$

Proof. Note that the assumption that $s_t^m > s_*^{c_1}$ for all t implies that s_t^m is strictly positive for all t. For any given t consider θ with $r^\theta \geq s_t^m$. For all c_i, $\eta(\{\theta \in \mu_{t+1}(c_i) | r^\theta \geq s_t^m\}) < k^{c_i}$. This is because, as in the proof of Claim 1, all such students θ face no peer costs, and as such, $\mu_{t+1}(\theta) = \mu_*(\theta)$. Because $s_{t+1}^m > s_*^{c_1}$ by assumption, no program c_i enrolls enough of these top students with scores $r^\theta \geq s_t^m$ at time $t+1$ to fill k^{c_i} measure of seats. Therefore, for any $c_i \in C$ if $\eta(\{\mu_{t+1}(c_i)\}) < k^{c_i}$ then $s_{t+1}(c_i) = 0$ and if $\eta(\{\mu_{t+1}(c_i)\}) \geq k^{c_i}$ then the score of the $(k^{c_i})^{th}$ highest scoring student is strictly less than s_t^m. □

As s_t^m, $t \geq 1$ is a strictly decreasing sequence and $s_t^m \in (s_*^{c_1}, 1]$, the sequence must converge to $S \geq s_*^{c_1}$. Suppose for contradiction that $S > s_*^{c_1}$. Let $M_{c_i}^{s_t}$ implicitly solve $k^{c_i} = \eta(\{\theta | r^\theta \geq s_t^m \text{ and } \mu_*(\theta) = c_i\}) + \eta(\{\theta | r^\theta \in [M_{c_i}^{s_t}, s_t^m)\})$ (if there is no such value, let $M_{c_i}^{s_t} = 0$), that is, $M_{c_i}^{s_t}$ would be the score of the k^{c_i}'th highest student enrolled at c_i in period $t+1$ if all students with scores above s_t^m attended their favorite program, and all of the students with scores below s_t^m most preferred program c_i. Recall that for all students θ with $r^{\theta} \geq s_t^m$, $\mu_*(\theta) = \mu_{t+1}(\theta)$. Therefore, $M_{c_i}^{s_t}$ is an upper bound on $s_{t+1}^{c_i}$. For any $s_t^m \geq S > s_*^{c_1}$, it must be that there is a unique $M_{c_i}^{s_t} < s_t^m$ for each c_i. Note also by assumption **A1**, it must be that $M_{c_i}^{s_t}$ is bounded away from s_t^m when $s_t^m > S > s_{c_1}^*$, i.e. there exists some $\delta > 0$ such that $s_t^m - M_{c_i}^{s_t} > \delta$ for all c_i if $s_t^m > S > s_{c_1}^*$. Therefore, for t such that $s_t^m - S < \delta$ (which must exist by the convergence hypothesis), $s_{t+1}^{c_i} < M_{c_i}^{s_t} < s_t^m - \delta < S$, which contradicts that s_t^m is a decreasing sequence that converges to S.

Suppose for contradiction that $S = s_*^{c_1}$. We know that $S > s_*^{c_j}$ for all $j \neq 1$. By a similar argument to the case in which $S > s_*^{c_1}$ we arrive at the conclusion that there exists t such that for all $t' > t$, $s_{t'}^{c_j} < S$ for all $j \neq 1$. Therefore, our contradiction hypothesis that for all t, $s_t^m > S$ is equivalent to the condition that for all $t' > t$, $s_{t'}^{c_1} > S$.

Consider any $t' > t$ and suppose $s_t^{c_1} > S = s_*^{c_1}$. Since $s_t^{c_j} < S = s_*^{c_1}$, we claim that $s_{t+1}^{c_1} < S$. To see this, note that any student θ with scores $r^{\theta} \in [S, 1]$ has $\mu_{t+1}(\theta) = c_1$ only if $\mu_*(\theta) = c_1$; these students face no peer costs from any program given μ_* and face a peer cost from c_1 given μ_t if $r^{\theta} \in [S, s_t^{c_1})$. Therefore, it must be that $s_{t+1}^{c_1} \leq S$, contradicting our assumption that for all $t' > t$, $s_t^{c_j} > S$. Claim 1 completes the proof of the **Base Case**.

□

Induction Step: Suppose $c_j \notin C'$ and that there exists some time t such that for all $t' \geq t$ all programs c_i, $i < j$ have $s_{t'}^{c_i} = s_*^{c_i}$ and $s_*^{c_i} \leq s_*^{c_{j-1}}$ for $i \geq j$. Then there exists $\bar{t}$ such that for all $t'' > \bar{t}$, $s_{t''}^{c_j} = s_*^{c_j}$ and $s_*^{c_j} > s_{t''}^{c_i}$ for all $i < j$.

Proof. The proof follows the case of the **Base Case**, and we therefore only summarize the arguments here. Take a program $c_j \in C \backslash \{c_0\}$:

1. If there is some time $\bar{t} - 1$ such that $s_{\bar{t}-1}^{c_i} \leq s_*^{c_j}$ for every c_i with $i \geq j$ and $s_{\bar{t}}^{c_k} = s_*^{c_k}$ for all c_k with $k < j$, then for all $t'' > \bar{t}$, $s_{t''}^{c_j} = s_*^{c_j}$ and $s_*^{c_j} > s_{t''}^{c_i}$ for all $i < j$.
2. Let $s_t^{m,j} = \max\limits_{c_i, i \geq j} s_t^{c_i}$. Then if $s_t^{m,j} > s_*^{c_j}$ for all t, $s_t^{m,j} > s_{t+1}^{m,j} > \ldots.$
3. There exists some $\bar{t} - 1$ such that $s_{\bar{t}-1}^{m,j} \leq s_*^{c_j}$.

The argument for the first claim is completely analogous to the one for the base case, with the change in notation of s_t^m to $s_t^{m,j} = \max\limits_{c_i, i \geq j} s_t^{c_i}$.

The second and third claims hold from the fact that the entire argument that we had before still stands after the programs with higher stable statistics have already converged. We can just use the same arguments with the preferences given the known statistics $s_T^{c_k} = s_*^{c_k}$ for $k < j$. □

We finish the proof by considering programs $c_j \in C'$, i.e. those for which $s_*^{c_j} = 0$. By our previous induction argument, there is some t such that for all $t' > t$, $s_{t'}^{c_i} = s_*^{c_i}$ and $s_{t'}^{c_j} < s_*^{c_i}$ for all $c_i \notin C'$ and all $c_j \in C'$. The following arguments hold for all programs $c_j \in C'$:

1. If there is some time $t-1$ such that $s_{t-1}^{c_j}=0$ for every $c_j \in C$ then $s_t^{c_j}=0$.
2. Let $s_t^{m,0}=\max_{c_j \in C'} s_t^{c_j}$. Then $s_t^{m,0} > s_{t+1}^{m,0} > ....$

To see that 1. holds, note that if $c_i \in C'$, $c_j \in C'$ for any $j > i$. Therefore all programs with high index have statistics lower or equal to 0, and then, by an argument analogous to the one before, they converge to their stable score right away. The argument for 2. is just analogous to the one from the previous case.

Therefore, it remains only to show that there exists some $\bar{t}-1$ such that $s_{\bar{t}-1}^{m,0}=0$. Given that $s_t^{m,0}$ is bounded and decreasing, it must converge. By a similar argument to above, we show that it cannot converge to $S > 0$. To show that $s_t^{m,0}$ cannot converge to 0 without ever reaching it in finite time, note that our genericity condition suggests that there is at most one program $c_J \in C'$ for which $\eta(\mu(c_J)) \geq k^{c_J}$, and for all other programs $c_j \in C'$, $\eta(\mu(c_j)) < k^{c_j}$. By assumption **A4** and our earlier arguments, for $\bar{t}-2$ such that $s_{\bar{t}-2}^{m,0}$ is sufficiently close to 0, $s_{t''}^{c_j}=0$ for all $t'' > \bar{t}-2$ and all $c_j \in C' \setminus \{c_J\}$. This implies that $s_{\bar{t}-1}^{c_J}=0$ since c_J is “less popular" at $\bar{t}-1$ than in the stable matching, which completes the argument. □

This proves our result for **Scenario 1**, in which there is no entry or exit. We will now extend our findings to the following scenario with entry and exit of programs.

Scenario 2: Not all programs are in block B_1, so that $B_1 \subsetneq C \setminus \{c_0\}$

Proof. We already know, from Proposition 2, that any sequence of stable matchings $\{\mu_t^*\}_{t\geq 1}$ for any Australian market are such that for all $c \in B_1$, $s^c(\lambda(\mu_t^*)) = s^c(\lambda(\mu_{t'}^*))$ for all $t' \geq 1$. What now needs to be shown, then, is that different stable scores for programs in B_2 do not change the stable scores of programs in B_1.

Note first that from bullet point 1. of Proposition 2, we know that for any $c_i \in B_1$ and $c_j \in B_2$ and a stable matching μ^*, $s^{c_i}(\lambda(\mu_t^*)) \geq s^{c_j}(\lambda(\mu_t^*))$.

Given that, to get the result we only need to follow the steps of Proposition 2 to see that each program in B_1 must have the same stable statistic and TIM will make the statistics reach them in finite time.

Take a s_0 vector of aggregate statistics. We will argue that s_2 is the same for any program in B_1 for any entry and exit realization at period $t=2$.

At period $t=1$ we get s_1. At period $t=2$ some programs in B_2 exit and enter the market, with some perceived statistics. Denote by s_t^{min} the lowest statistics value among programs in B_1 at time t and by s_*^N the lowest stable statistic for any program in B_1. From the previous point we know that such value exists, as all programs in B_1 have the same stable statistics at any stable matching. The argument will come down to two simple points:

(a) If $s_t^{min} \leq s_*^N$ at period t, $\mu^{t'}(\theta) \in B_1$ for any $t' > t$ and any student-type θ with $r^\theta \geq s_*^N$.

(b) We have that $s_t^{min} \leq s_*^N$ for $t = 1$.

To see that point (a) holds, note that any such student-type θ with $r^\theta \geq s_*^N$ will, at time $t+1$, rank all programs in B_1 with $r^\theta \geq s_t^{c_i}$ higher than any program in B_2. This student must be able to enroll at some program in B_1, otherwise s^* would not be a stable aggregate statistics vector. To see this more clearly, just note that at any stable matching, all students with scores $r^\theta \geq s_*^N$ go to some program in B_1. Therefore for a students not to be able to get into any B_1 program at $t+1$, we would need to $p_t^{min} > r^\theta \geq s_*^N$, which is clearly not possible, as p_t^{min} must be lower or equal to s_*^N.

For point (b), note first that if $s_t^{min} \leq s_*^N$ at either $t = 0$ or $t = 1$, we are done, by the previous point. Suppose, then, that, $s_t^{min} > s_*^N$ at $t = 0$ and $t = 1$. If $s_t^{min} > s_*^N$ at $t = 1$ as well, then there must be a set of students with $r^\theta \geq s_*^N$ who did not enroll at any program in B_1 at $t = 1$. But then the lack of such students in B_1 programs must make at least one of them have a statistic lower than s_*^N, and therefore $s_t^{min} \leq s_*^N$ at $t = 1$, a contradiction.

To conclude the proof, note that the market for programs in B_1 and student-types θ with $r^\theta \geq s_*^N$ from $t = 2$ onward is unaffected by entry and exit realizations of programs is B_2. Therefore the convergence of statistics of programs in B_1 happens exactly as in Scenario 1.

We have then that entry and exit of programs of programs widely seen as less desirable (those in block B_2) does not affect convergence of statistics of ones seen as more desirable (B_1 programs), as long as we have block-correlated student preferences. □

Remark 5

Proof. We verify each of the desired conditions separately.

A2 This follows from **AA6** when $C = B_1 \cup \{c_0\}$.

A5 This follows from **AA2** and the construction of $s(\cdot)$.

A6 This follows from **A1** and the continuity of $f^{\theta,c}(\cdot,\cdot)$ in its second argument for each $\theta \in \Theta$ and $c \in C$ (see **AA2**).

A7 Note that if at some $\mu \in M$ and $\epsilon > 0$ it is the case that $\eta(\mu(c)) < q^c$, then $s^c(\mu) = 0$, and there exists $\delta > 0$ such that if $\nu \in M$ satisfies $||\lambda(\mu) - \lambda(\nu)||_\infty < \delta$, then $s^c(\nu) = 0$. Therefore, we focus on the case in which $\eta(\mu(c)) \geq q^c$.

We say that the set of students $\mu(c)$ has a *hole* if there is an interval of scores $(r^l, r^h) \subset (0,1)$, with $r^h > r^l$, such that $\eta(\theta | \{r^{\theta,c} \leq r^l\} \cap \{\mu(\theta) = c\}) > 0$ but $\eta(\theta | \{r^{\theta,c} \in (r^l, r^h)\} \cap \{\mu(\theta) = c\}) = 0$. In words, there is an interval of scores in which a zero measure of students are assigned to a program, even though there is a positive measure of students enrolled at the program with scores below the interval. If a matching μ is such that for all c, $\mu(c)$ has no holes, then we say that μ is *score connected*.

The proof is in two steps: 1) All $\mu \in M$ are score connected (i.e. μ is score connected if it is a market clearing matching), and 2) If all $\mu \in M$ are score connected, then $s(\cdot)$ satisfies **A7**.

Step 1: All $\mu \in M$ are score connected.

Let $\mu \in M$. Then $\mu = A(p, \lambda)$ for some (p, λ). Suppose for contradiction that there is a hole (r^l, r^h) at $\mu(c)$ for some $c \in C$. Then there exists a positive-measure set of student types Θ' such that $r^{\theta',c} \leq r^l$ and $\mu(\theta') = c$ for all $\theta' \in \Theta'$. This implies that the cutoff p^c is such that $p^c < r^l$. By **A2** (which follows from **AA6**, as shown earlier in the proof of this Remark), there exists a positive-measure set of students $\hat{\Theta}$ such that $r^{\hat{\theta},c} \in (r^l, r^h)$ for all $\hat{\theta} \in \hat{\Theta}$ who strictly prefer c to any other program at μ. This contradicts that $\mu \in M$ as $\mu(\hat{\theta}) \neq c = D^{\hat{\theta}}(p, \lambda)$ for all $\hat{\theta} \in \hat{\Theta}$.

Step 2: If all $\mu \in M$ are score connected, then $s(\cdot)$ satisfies **A7**.

Recall that for all $c \in C$, $k^c \in [0,1]$ is such that $s^c(\lambda(\mu))$ equals the supremum value of r^θ for which $\eta(\{\theta' \in \mu(c) | r^{\theta'} > r^\theta\}) = k^c$ (if such a number exists, and 0 otherwise).

Fix $c \in C$, $\epsilon > 0$, and $\mu \in M$. Because μ is score connected, $\mu(c)$ has no holes and so there is a unique value $s^c(\lambda(\mu)) \in [0,1]$ that satisfies $\eta(\{\theta' \in \mu(c) | r^{\theta'} > s^c(\lambda(\mu))\}) = k^c$. Notice also that $s^c(\lambda(\mu))$ is continuous in k^c. Take $\delta > 0$ and any $\nu \in M$ such that $\eta(\{\mu(c) \setminus (\mu(c) \cap \nu(c)\}) < \delta$ and $\eta(\{\nu(c) \setminus (\mu(c) \cap \nu(c)\}) < \delta$. It suffices to show that as $\delta \to 0$, $s^c(\lambda(\nu)) \to s^c(\lambda(\mu))$. Let $s^c(\lambda(\mu), \delta)$ be defined implicitly by $\eta(\{\theta' \in \mu(c) | r^{\theta'} > s^c(\lambda(\mu), \delta)\}) = k^c + \delta$ and $s^c(\lambda(\mu), -\delta)$ be defined implicitly by $\eta(\{\theta' \in \mu(c) | r^{\theta'} > s^c(\lambda(\mu), -\delta)\}) = k^c - \delta$. Then because ν is score connected (as $\nu \in M$), it must be the case that $s^c(\lambda(\nu)) \in (s^c(\lambda(\mu), -\delta), s^c(\lambda(\mu), \delta))$. By the continuity of $s^c(\lambda(\mu))$ in k^c, $s^c(\lambda(\mu), \delta) \overset{\delta \to 0}{\to} s^c(\lambda(\mu))$ and $s^c(\lambda(\mu), -\delta) \overset{\delta \to 0}{\to} s^c(\lambda(\mu))$. Therefore, $s^c(\lambda(\nu)) \to s^c(\lambda(\mu))$ as $\delta \to 0$.

□

Remark 6

Proof. Let $\epsilon \in (0,1)$ and let E^ϵ be an Australian market where $1-\epsilon$ measure of students have common intrinsic preferences, that is $\eta\{\theta | v^{c_1,\theta} > v^{c_2,\theta} > ... > v^{c_N,\theta}\} = 1-\epsilon$. Let $\tilde{E}^\epsilon$ be a market that differs from E^ϵ only in that we permute student intrinsic preference such that $v^{c_1,\theta} > v^{c_2,\theta} > ... > v^{c_N,\theta}$ for almost all θ. Let $\tilde{\mu}_*$ and μ_* represent the unique stable matchings in $\tilde{E}^\epsilon$ and E^ϵ, respectively. Let $\tilde{s}_*$ and s_* represent the vector of k^{th} highest scores at each program in stable matchings $\tilde{\mu}_*$ and μ_*, respectively. The following steps together prove our desired result.

Step 1: For any $\delta > 0$ there exists $\epsilon' > 0$ such that for all $\epsilon < \epsilon'$, $||s_* - \tilde{s}_*||_\infty < \delta$.

Step 2: For any $i \neq j$ such that $\tilde{s}_*^{c_i} > 0$, there exists $\delta > 0$ such that $|\tilde{s}_*^{c_i} - \tilde{s}_*^{c_j}| > \delta$. Similarly, there exists ϵ' such that for any $\epsilon < \epsilon'$, $|s_*^{c_i} - s_*^{c_j}| > \delta$.

Step 3: Given any μ_0, $\tilde{s}_{N+1} = \tilde{s}_*$ in the TIM process in market $\tilde{E}^\epsilon$.

Step 4: For any $\delta > 0$ and μ_0, there exists $\epsilon' > 0$ such that for $\epsilon < \epsilon'$, $||\tilde{s}_t - s_t||_\infty < \delta$ for all $t \leq 3N+1$.

Step 5: For any μ_0 there exists $\epsilon' > 0$ such that for any $\epsilon < \epsilon'$, the TIM process in E^ϵ yields s_{N+1} such that $||s_{N+1} - s_*||_\infty < \delta$. Continuing on in the TIM process, $s_{3N+1} = s_*$.

We now prove each step in the order presented:

Step 1: For any $\delta > 0$ there exists $\epsilon' > 0$ such that for all $\epsilon < \epsilon'$, $||s_* - \tilde{s}_*||_\infty < \delta$.

Proof. The statement follows from the pseudo-serial dictatorship mechanism presented in Proposition 2. By construction, $\tilde{s}_*^{c_1} \geq \tilde{s}_*^{c_2} \geq ... \geq \tilde{s}_*^{c_N}$. For any $\gamma_1 > 0$, exists ϵ_1 such that for all $\epsilon < \epsilon_1$, $1-\gamma_1$ measure of students attend the same program in the first step of the mechanism in economies E^ϵ and $\tilde{E}^\epsilon$, respectively, by Lemma B.3 of Azevedo and Leshno (2016). Remark 5 finds that μ_1 is score connected and therefore (following Step 2 of that remark), for sufficiently small γ_1, $|s_1^{c_1} - \tilde{s}_1^{c_1}| < \delta$, where $s_1^{c_1}$ and $\tilde{s}_1^{c_1}$ are the summary statistic of program c_1 in the first stage of the mechanism in economies E^ϵ and $\tilde{E}^\epsilon$, respectively. By Proposition 2, it is the case that $s_1^{c_1} = s_*^{c_1}$ and $\tilde{s}_1^{c_1} = \tilde{s}_*^{c_1}$. By induction it follows by this argument that there exists $\epsilon_i > 0$ such that for all $\epsilon < \epsilon_i$, $|s_*^{c_i} - \tilde{s}_*^{c_i}| < \delta$. Then, take $\epsilon' = \min\{\epsilon_1, ..., \epsilon_N\}$ to complete the claim. □

Step 2: For any $i \neq j$ such that $\tilde{s}_*^{c_i} > 0$, there exists $\delta > 0$ such that $|\tilde{s}_*^{c_i} - \tilde{s}_*^{c_j}| > \delta$. Similarly, there exists ϵ' such that for any $\epsilon < \epsilon'$, $|s_*^{c_i} - s_*^{c_j}| > \delta$.

Proof. If $\tilde{s}_*^{c_i} > 0$ then $\tilde{s}_*^{c_j} > \tilde{s}_*^{c_i}$ for $j < i$. This follows because $v^{c_j,\theta} > v^{c_i,\theta}$ for almost all θ, and therefore at most a zero-measure set of students with scores $r^\theta \geq s_*^{c_j}$ can be matched to c_i, $\eta(\{\theta \in$

$\tilde{\mu}_*(c_i)|r^\theta \geq s_*^{c_j}\}) = 0$. By similar logic, $\tilde{s}_*^{c_j} < \tilde{s}_*^{c_i}$ if $j > i$. Let N' represent the subset of programs such that $\tilde{s}_*^{c_i} > 0$ for all $i \in N'$. Therefore, any $\delta \in (0, \min_{i \in N'} s_*^{c_i} - s_*^{c_{i+1}}]$ satisfies our requirement.

That there exists ϵ' such that for any $\epsilon < \epsilon'$, $s_*^{c_i} - s_*^{c_{i+1}} > \delta$ for all $i \in N'$ follows from the previous argument and the conclusion of Step 1.

□

Step 3: Given any μ_0, $\tilde{s}_{N+1} = \tilde{s}_*$ in the TIM process in market $\tilde{E}^\epsilon$.

Proof. Fix $t > 0$, and suppose that all c_j with $j < i \leq N$ are such that $\tilde{s}_t^{c_j} = \tilde{s}_*^{c_j}$. If $\tilde{s}_t^{c_i} \leq \tilde{s}_*^{c_i}$ then $\tilde{s}_{t'}^{c_i} = \tilde{s}_*^{c_i}$ for all $t' \geq t+1$. To see this, consider the set of students $\{\theta | r^\theta \geq \tilde{s}_*^{c_i}\}$. All such students θ will have $\tilde{\mu}_{t+1}(\theta) = c_k$ for $k \leq i$ by Assumption **AA2** and the fact that intrinsic preferences are fully aligned in market $\tilde{E}^\epsilon$. Therefore, $\tilde{s}_t^{c_k} \leq \tilde{s}_*^{c_i}$ for all $k > i$. By **Scenario 1** from the proof of Theorem 3, this implies that $\tilde{s}_{t'}^{c_i} = \tilde{s}_*^{c_i}$ for all $t' \geq t+1$.

The following induction argument shows that $\tilde{s}_{N+1} = \tilde{s}_*$.

Base Case: $\tilde{s}_2^{c_1} = \tilde{s}_*^{c_1}$ and $\tilde{s}_2^{c_2} \leq \tilde{s}_*^{c_2}$.

By the fact that intrinsic preferences are fully aligned, it is the case that $\tilde{s}_*^{c_1} = \max\{1 - k^{c_1}, 0\}$. Therefore, for any μ_0, $\tilde{s}_1^{c_1} \leq \tilde{s}_*^{c_1}$. We then have that almost every student $\theta \in \{\theta | r^\theta \geq \tilde{s}_*^{c_1}\}$ will have $\mu_2(\theta) = c_1$. Moreover, because $s_1^{c_1} \leq s_*^{c_1}$ and students have big-fish preferences (Assumption **AA2**), it must be the case that $\eta(\{\theta | \tilde{\mu}_2(\theta) = c_1\} \cap \{\theta | r^\theta \geq \tilde{s}_*^{c_2}\}) \geq \eta(\{\theta | \tilde{\mu}_*(\theta) = c_1\} \cap \{\theta | r^\theta \geq \tilde{s}_*^{c_2}\})$. As a result, $\tilde{s}_2^{c_2} \leq \tilde{s}_*^{c_2}$.

Induction Case: If at time period $t > 1$ it is the case that $\tilde{s}_t^{c_j} = \tilde{s}_*^{c_j}$ for all $j < i \leq N$ (if there exists $0 < j < i$) and $\tilde{s}_t^{c_i} \leq \tilde{s}_*^{c_i}$ then $\tilde{s}_{t+1}^{c_i} = \tilde{s}_*^{c_i}$ and $\tilde{s}_{t+1}^{c_{i+1}} \leq \tilde{s}_*^{c_{i+1}}$ if $i+1 \leq N$.

It remains only to show that if $i+1 \leq N$, then $\tilde{s}_{t+1}^{c_{i+1}} \leq \tilde{s}_*^{c_{i+1}}$. This follows a similar logic as in the base case; $\eta(\{\theta | \tilde{\mu}_{t+1}(\theta) = c_k, k < i+1\} \cap \{\theta | r^\theta \geq \tilde{s}_*^{c_{i+1}}\}) \geq \eta(\{\theta | \tilde{\mu}_*(\theta) = c_k, k < i+1\} \cap \{\theta | r^\theta \geq \tilde{s}_*^{c_{i+1}}\})$. As a result, $\tilde{s}_{t+1}^{c_{i+1}} \leq \tilde{s}_*^{c_{i+1}}$.

□

Step 4: For any $\delta > 0$ and μ_0, there exists $\epsilon' > 0$ such that for $\epsilon < \epsilon'$, $||\tilde{s}_t - s_t||_\infty < \delta$ for all $t \leq 3N+1$.

Proof. Fix μ_0 and $\delta > 0$. Define μ_t and $\tilde{\mu}_t$ as the matchings formed at t for economies E^ϵ and $\tilde{E}^\epsilon$, respectively. By Lemma B.3 of Azevedo and Leshno (2016) for any $\gamma_1 > 0$ there exists ϵ_1 such that for all $\epsilon < \epsilon_1$, $\eta(\{\theta | \mu_1(\theta) \neq \tilde{\mu}_1(\theta)\} < \gamma_1$. Assumption **A2** implies that for sufficiently small γ_1, $||s_1 - \tilde{s}_1||_\infty < \delta$. By repeated application of Remark 2 and Lemma B.3 of Azevedo and Leshno (2016), there exists ϵ_t such that for all $\epsilon < \epsilon_t$, $\eta(\{\theta | \mu_t(\theta) \neq \tilde{\mu}_t(\theta)\} < \gamma_t$. Assumption **A2** implies that for sufficiently small γ_t, $||s_t - \tilde{s}_t||_\infty < \delta$. To complete the result, let $\epsilon' = \min_{t \leq 3N+1} \epsilon_t$.

□

Step 5: For any μ_0 there exists $\epsilon' > 0$ such that for any $\epsilon < \epsilon'$, the TIM process in E^ϵ yields s_{N+1} such that $||s_{N+1} - s_*||_\infty < \delta$. Continuing on in the TIM process, $s_{3N+1} = s_*$.

Proof. The first statement holds by the results of Steps 1, 3, and 4.

Again letting N' represent the subset of programs such that $\tilde{s}_*^{c_i} > 0$ for all $i \in N'$, Steps 2 and 4 imply that for sufficiently small ϵ, $s_t^{c_i} - s_t^{c_{i+1}} > \delta$ for all $i \in N'$ and all $t \in \{N+1, ..., 3N+1\}$.

Therefore, it remains only to show that $s_{3N+1} = s_*$. By the argument in the last paragraph, we know that either $s_{N+1}^{c_1} > s_{N+1}^{c_j}$ for all $j \neq 1$ or $s_{N+1} = s_* = \{0, 0, ..., 0\}$. If $s_t^{c_1} \leq s_*^{c_1}$, we have that $s_{t+1}^{c_1} = s_*^{c_1}$, by the proof of Theorem 3. If $s_t^{c_1} > s_*^{c_1}$, we have that $s_{t+1}^{c_1} \leq s_*^{c_1}$. By Steps 1-4, it must be the case that $s_{N+1}^{c_1} > s_*^{c_1} > s_{N+1}^{c_j}$ for all $j \neq 1$ for sufficiently small ϵ. By Assumption **AA2**, it must be that $\eta(\{\theta | c_1 \succ^{\theta | s_{N+1}} c_j \text{ for all } j \neq 1 \text{ AND } r^\theta > s_*^{c_1}\}) \leq k^{c_1}$. By the argument presented before, this means that $s_{t+2}^{c_1} = s_*^{c_1}$. The argument for the other programs hold analogously, with each program c_i reaching its steady-state summary statistic at most two periods after program c_{i-1}.

□

□

Example With Tight Bound of $N+1$ Periods For Convergence

Let E be an Australian market in which $k^{c_i} < q^{c_i}$ for each $c_i \in C$ and in which is an undersupply of seats: $\sum_i q^{c_i} < 1$.

Programs are almost universally ranked by students and more popular programs are more "competitive": $\eta\{\theta | v^{c_1,\theta} > v^{c_2,\theta} > ... > v^{c_N,\theta}\} = 1 - \epsilon$ for some small ϵ and $k_{c_i} < k_{c_j}$ for $0 < i < j$. Moreover, peer preferences are strong: $f^\theta(r^\theta, s^{c_i}) > v^{c_1}$ whenever $r^\theta < s^{c_i} - \epsilon$.

For sufficiently small ϵ, it is the case that $s_*^{c_1} > s_*^{c_2} > ... > s_*^{c_N} > 0$. We show that for a given starting condition μ_0 and sufficiently small ϵ, the market does not (approximately) converge in strictly fewer than $N+1$ periods.

Let μ_0 be such that $s_0^{c_i} > s_*^{c_1} + \epsilon$ for all i. Then by our assumption on peer preferences, and our assumption that $k_{c_i} < k_{c_j}$ for $0 < i < j$, no program c_i fills k^{c_i} seats at $t = 1$, $\eta(\mu_1(c_i)) < k^{c_i}$. Therefore, $s_1^{c_i} = 0$ for all $c_i \in C$.

At $t = 1$, all students of sufficiently high score attend their stable partner for sufficiently small ϵ: $\mu_1(\theta) = \mu_*(\theta)$for all $\theta \in \{\theta | r^\theta > s_*^{c_1}\}$. This follows because that students face no peer costs at any program due to $s_1^{c_i} = 0$ for all $c_i \in C$. However, by Steps 3 and 4 of the proof of Remark 6, $s_2^{c_2} < s_*^{c_2}$. As a result, $s_t^{c_2}$ does not reach steady state until $t = 3$.

We can continue this argument to show that for each $0 < t \leq N$, $s_t^{c_t} < s_*^{c_t}$, which implies that $s_t = s_{N+1}$ only for $t \geq N+1$.

A.2 Additional evidence, figures, and tables

Table A.1: Summary Statistics of Adjustments to pre ROL

Variable	Obs	Mean	Std. Dev.	P25	P50	P75
Only Switchers	167352	.12	.32			
Only Adders	167352	.07	.25			
Only Removers	167352	.03	.16			
Any switch	167352	.23	.42			
Any add	167352	.27	.44			
Any remove	167352	.2	.4			
Any change	167352	.43	.5			
Nr. of switches	167352	.7	1.8	0	0	0
Nr. of adds	167352	.63	1.32	0	0	1
Nr. of removes	167352	.47	1.18	0	0	0
Nr. of changes	167352	1.81	2.85	0	0	3
Share of final list switched	167352	.13	.25	0	0	0.2
Share of final list added	167352	.09	.18	0	0	.13
Share of final list removed	167352	.06	.14	0	0	0
Share of final list changed	167352	.28	.36	0	0	.6

This table summarizes adjustments students make to their submitted ROLs once they learn their final ATAR score. Rows denoted by "Only..." present the share of students who conduct only the stated adjustment to their pre ROL. Rows denoted by "Any..." present the share of students who conduct the stated adjustment to their pre ROL. Rows denoted by "Nr...." present the average number of the stated adjustments to the pre ROL across students. Rows denoted by "Share..." present the average across students of the ratio of the number of the stated adjustments made to the length of the pre ROL. We use the pre- and post-ROL sample from 2010-2016.

Table A.2: Across Time Applicant Response to Program PYS, including CYS and Lagged PYS Values

	(1)	(2)	(3)	(4)	(5)
	Avg. Applicant Score	# of Applicants	% of Applicants	% of Applicants Higher Score	% of Applicants Lower Score
Current Statistic	0.085***	0.584*	0.003**	0.007***	0.003
	(0.018)	(0.257)	(0.001)	(0.001)	(0.001)
Past Year Statistic	0.265***	-2.863***	-0.011***	-0.010***	-0.019***
	(0.018)	(0.288)	(0.001)	(0.002)	(0.002)
2 Years Ago Statistic	0.035*	-0.198	-0.000	0.003*	0.001
	(0.015)	(0.193)	(0.001)	(0.001)	(0.001)
3 Years Ago Statistic	0.075***	-0.559*	0.001	0.003**	-0.002
	(0.015)	(0.241)	(0.001)	(0.001)	(0.001)
Observations	8,582	8,582	8,582	8,582	8,582

This table shows the estimated β coefficients of a regression similar to (1) where we additionally include the Current Year Statistic and the 2 and 3 Years Ago Statistic of the program. $y_{c,t}$ is the average applicant score, number of students who apply, percent of students who apply, or the percent of students who apply with ATAR scores higher/lower than the program PYS for program c in year t. Standard errors in parentheses, clustered at program level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.3 shows the impact of the PYS on applicant demand for very young (2 years old) versus very old programs (14 or more years old) in our sample. Specifically, we estimate the following regression:

$$\begin{aligned} y_{c,t} = {} & \beta PYS_{c,t} + \gamma Age_{c,t} + \lambda Age\ Known_c + \delta_0 PYS_{c,t} \times Age_{c,t} + \delta_1 PYS_{c,t} \times Age\ Known_c \\ & + \delta_2 Age_{c,t} \times Age\ Known_c + \delta_3 PYS_{c,t} \times Age_{c,t} \times Age\ Known_c + \alpha_c + \alpha_t + \epsilon_{c,t} \end{aligned} \tag{A.2}$$

where $y_{c,t}$ denotes the average applicant score, number of students who apply, percent of students who apply, or the percent of students who apply with ATAR scores higher/lower than the program PYS for program c in year t. $Age_{c,t}$ is the number of years we observe a program in the sample and $Age\ Known_c$ is a dummy that is equal to one if the program is established within our sample period and we can thus be certain of its age. We again include year and program fixed effects (α_c and α_t, respectively) to isolate variation in PYS that is happening within program over time. Table A.3 presents the linear combination of coefficients of Equation A.2 for (i) 2 year old program where we know the true age, i.e. that start within our sample period ($Age\ Known_c$=1) and (ii) programs that we observe for every year in the sample, i.e. programs that are 14 or more years in existence ($Age\ Known_c$=0).

We note that the outcomes in Columns 2, 3, and 5 are nearly identical between the oldest and newest programs and the effect in Column 1 is larger for the oldest programs, which we would not expect if students cared about their fit with a program and not their peers. One anomalous

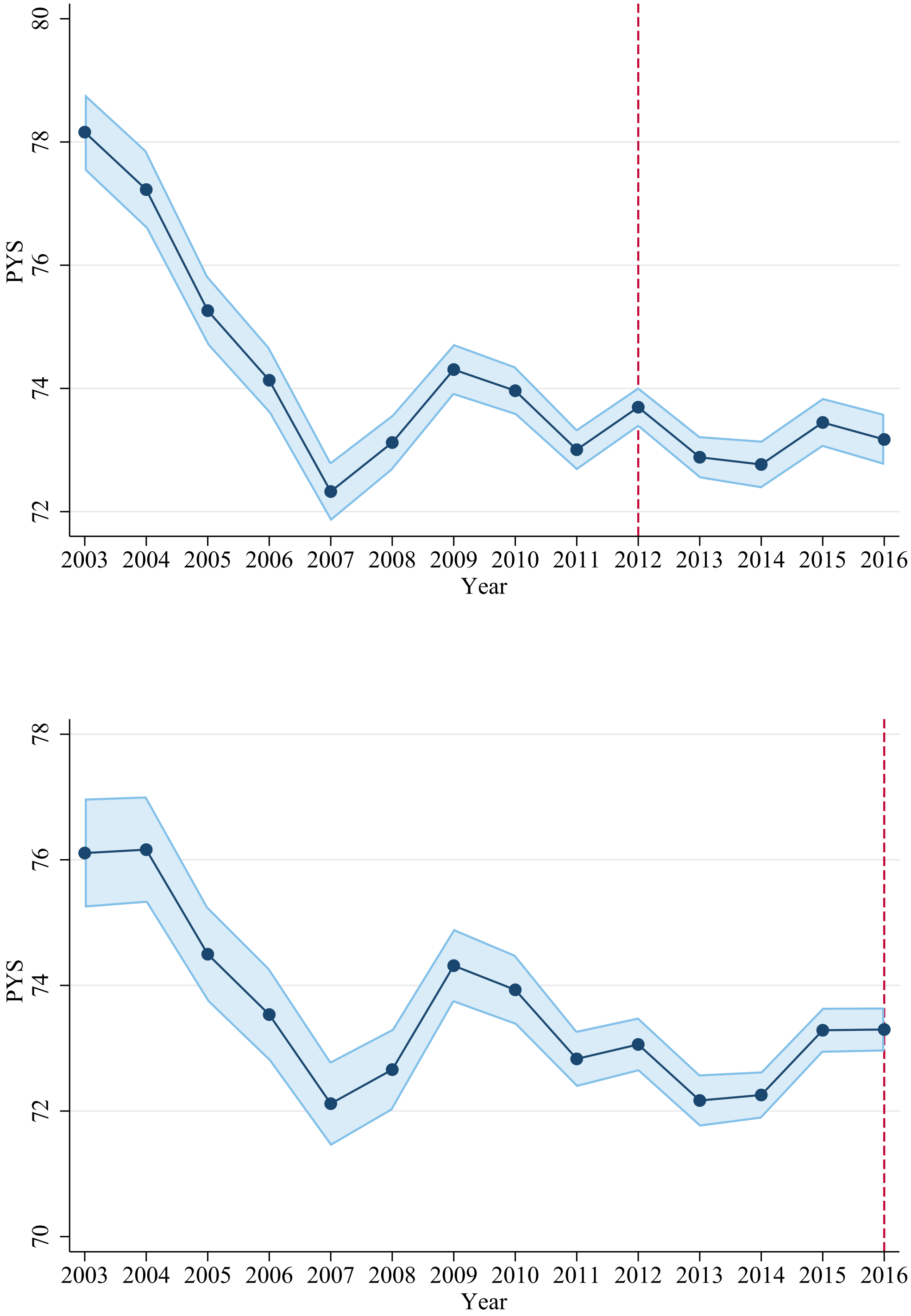
Table A.3: Impact of PYS on Applicant Demand for New and Established Programs

	(1) Avg. Applicant Score	(2) # of Applicants	(3) % of Applicants	(4) % of Applicants Higher Score	(5) % of Applicants Lower Score
2 Year Old Programs	0.274 ***	-2.002 ***	-0.007 ***	-0.003	-0.014 ***
	(0.023)	(0.197)	(0.001)	(.002)	(0.001)
14+ Year Old Programs	0.367 ***	-1.867 ***	-0.008 ***	-0.039 *	-0.015 ***
	(0.029)	(0.657)	(0.002)	(0.022)	(0.003)
Observations	14,850	14,850	14,850	14,850	14,850

This table shows the linear combination of estimated coefficients for Equation A.2 for (i) 2 year old programs (ii) 14 or more year old programs. For 2 year old programs we restrict on programs that start within our sample period where we can thus be certain of their true age (*Age Known*$_c$=1). For 14 or more year old programs we restrict on those programs that are already in existence when our sample starts and that we then observe for every consecutive year in our sample, meaning they will be at least 14 or more years old (*Age Known*$_c$=0). Standard errors in parentheses, clustered at the program level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

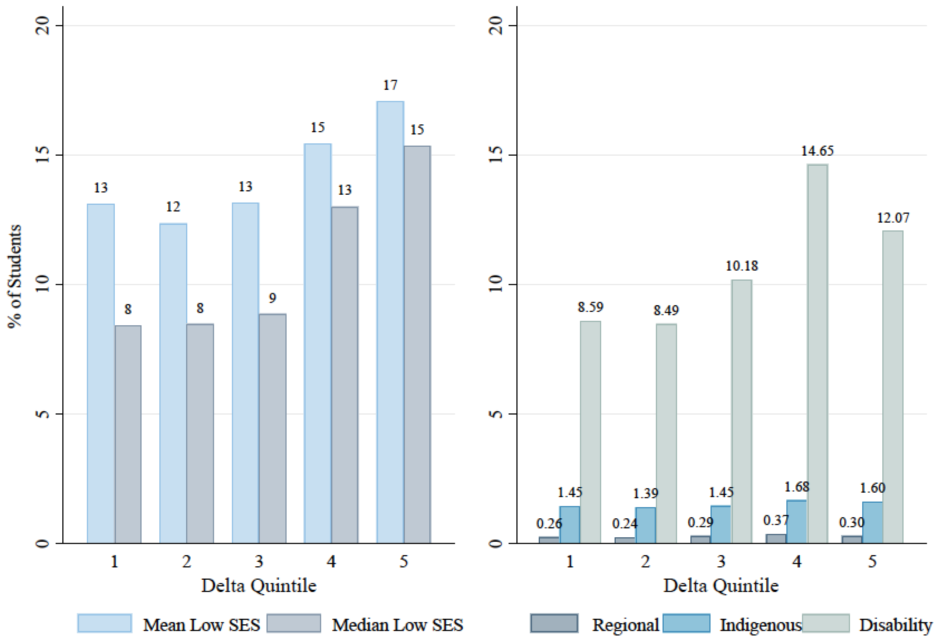
finding is the negative, but noisy coefficient in Column 4 for the oldest programs. This is at least in part due to a small sample issue (as we discuss later in Figure 8, the oldest programs typically have the highest PYS value, implying that a small fraction of students have ATAR scores above the PYS of these programs). This is supported by the nearly-identical point estimates in Columns 3 and 5 across the oldest and youngest programs (i.e. holding the point estimates in Columns 4 and 5 fixed, if there were a significant fraction of students with ATAR scores above the PYS of the oldest programs, we would mechanically expect a more negative coefficient in Column 3). Finally, we note that the coefficient in Column 4 for programs that are known to be exactly 13 years old is statistically indistinguishable from 0.

Figure A.1: Convergence test for matching in 2012 and 2016



The top panel groups together programs that have a similar PYS (within a 10 point band of 70) in 2012. It then follows the group's **average** PYS both forward and backwards in time. It shows that programs with similar PYSs in 2012 have converged from a more disperse distribution over time, and continue to converge even after 2012. The bottom figure repeats the same exercise, instead grouping together programs with a similar PYS in 2016.

Figure A.2: Demographics and Δ



Notes