DISCUSSION PAPER SERIES

DP16544
(v. 3)

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

Natalie Cox, Bobby Pakzad-Hurson and Ricardo Fonseca

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

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

Natalie Cox, Bobby Pakzad-Hurson and Ricardo Fonseca

Discussion Paper DP16544
First Published 13 September 2021
This Revision 24 May 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Organizational Economics

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

Abstract

Can a centralized school choice clearinghouse generate a stable matching if it does not allow students to express their preferences over both programs and 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 induces a tâtonnement process wherein the distributions of former student types play the role of prices. We theoretically model this process and develop a test for match stability. We implement this test empirically in the college admissions market in New SouthWales (NSW), Australia, where we find evidence of preferences over relative peer ability. We show that the NSW market fails to converge to stability over time. We propose a new mechanism that improves upon the current design, and we show that this mechanism generates a stable matching in the NSW market.

JEL Classification: I21, D47, C78

Keywords: market design, College Admissions, Peer Preferences

Natalie Cox - nbachas@princeton.edu
Princeton University and CEPR

Bobby Pakzad-Hurson - bph@brown.edu
Brown University

Ricardo Fonseca - ricardo_fonseca@brown.edu
Brown University

Acknowledgements
For very helpful comments and discussions we thank Eduardo Azevedo, YingHua He, Fuhito Kojima, Maciej Kotowski, Jacob Leshno, Shengwu Li, Margaux Luflade, George Mailath, Matt Pecenco, Tayfun Sönmez, Ran Shorrer, Utku Ünver, Rakesh Vohra, Bumin Yenmez, and seminar audience members at Boston College, CMU/Pitt, Brown, Penn, MSR New England, ITAM, Matching in Practice Workshop ’21, EAMMO ’21, and NBER Summer Institute (Education). We are grateful to Joanna Tasmin and especially Clemens Lehner for excellent research assistance.

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

Natalie Cox† Ricardo Fonseca‡ Bobak Pakzad-Hurson§

May 24, 2022

Abstract

Can a school-choice clearinghouse generate a stable matching if it does not allow students to express preferences over peers? Theoretically, we show stable matchings exist with peer preferences under mild conditions but finding one via canonical mechanisms is unlikely. Increasing transparency about the previous cohort's matching induces a tâtonnement process wherein prior matchings function as prices. We develop a test for stability and implement it empirically in the college admissions market in New South Wales, Australia. We find evidence of preferences over *relative* peer ability, but no convergence to stability. We propose a mechanism improving upon the current assignment process.

*For very helpful comments and discussions we thank Mohammad Akbarpour, Eduardo Azevedo, Yan Chen, YingHua He, Fuhito Kojima, Maciej Kotowski, Jacob Leshno, Shengwu Li, Margaux Luflade, George Mailath, Ellen Muir, Matt Pecenco, Paul Milgrom, Alex Rees-Jones, Al Roth, Tayfun Sönmez, Ran Shorrer, Utku Ünver, Rakesh Vohra, Bumin Yenmez, and seminar audience members at Boston College, Brown, CMU/Pitt, Penn, MSR New England, ITAM, Stanford, UCSB, Wharton, Matching in Practice Workshop '21, EAMMO '21, and 2021 NBER Summer Institute (Education). We are grateful to Joanna Tasmin and especially to Clemens Lehner for excellent research assistance.

An extended abstract of this paper appears at EC '22.

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

‡Brown University, 57 Waterman Street, Providence, RI 02906. Email: ricardo_fonseca@brown.edu

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

I Introduction

Following the application of matching theory to education markets (Abdulkadiroğlu and Sönmez, 2003), centralized mechanisms are now commonly used to allocate seats at schools and colleges in at least 46 countries (Neilson, 2019). Creating a stable matching–one in which no agent wants to deviate and rematch with a willing partner (or remain unmatched)–is often viewed as a chief concern in these settings (Roth, 2002). Student preferences over educational programs may depend on a variety of factors, including both *exogenous* characteristics such as location, and the *endogenous* characteristics of peers in their cohort. However, matching mechanisms used in school choice markets only ensure a stable matching if student preferences do not depend on their peers.

In this paper, we seek to answer three questions, which we believe must be investigated simultaneously to understand the role of peer preferences in present-day school choice markets: Do students have peer preferences, and do their preferences depend on relative peer ability? Do stable matchings exist when students have peer preferences? What are the consequences of failing to account for peer preferences in a centralized matching mechanism–specifically, what happens if a "stable" matching mechanism is misspecified, in that students are only allowed to express preferences over educational programs, but have preferences over both programs and their peers?

Empirically, we use data from the centralized matching market for college admissions in New South Wales (NSW), Australia where, by law, programs must provide applicants with information about the ability distribution of the existing student body. We show that students' ordinal rankings over programs are affected by information about potential peers. Our data suggest that students typically prefer not to match with a program where they are near the bottom of the ability distribution. This pattern accords with the "big-fish-little-pond effect," which has been well documented in the education literature, wherein a student being lower in the "ability" distribution leads to psychic costs (Marsh et al., 2008; Seaton, Marsh, and Craven, 2009; Pop-Eleches and Urquiola, 2013) and declines in achievement (Carrell, Sacerdote, and West, 2013).[1] These preferences are also reflected in the common practice of "redshirting" kindergartners; parents delay entry for an additional year so their child can be among the oldest and most developed in the class (Dhuey et al., 2019).[2]

Theoretically, we study large matching markets and find that a stable matching exists under mild

[1] 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); 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. See Sacerdote (2011) for a literature review of other peer effects. At the college level, there is evidence of peer effects in college student achievement. Sacerdote (2001) and Stinebrickner and Stinebrickner (2006) 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.

[2] The original use of the term "redshirting" stems from college athletes who delay participating in athletics for a year, and which can reflect a similar preference for developmental advantage over peers.

conditions when students have peer preferences. However, mechanisms used in practice that do not solicit student preferences over peers are unlikely to yield a stable matching. Over time, we show that the market may or may not converge to a stable matching.

We combine theory and empirics to demonstrate the risks of failing to explicitly account for peer preferences in school choice markets. Our model derives a simple test for convergence to a stable matching. Using data on admissions outcomes in NSW over 14 years, we verify our theoretical prediction that the top "ability" part of the market converges while the bottom does not. We propose a new mechanism, and prove that it generates an approximately stable matching in the NSW market.

While a market designer may have specific desires (eg. to maximize value added), an axiomatic characterization informs how to account for peer preferences in centralized school choice markets. Three desirable matching properties: individual rationality, non-wastefulness, and fairness (Balinski and Sönmez, 1999)[3] are jointly identical in our setting to (pairwise) stability *taking into account students' preferences over programs given the distribution of peers.* This characterization leads us to take a positive rather than normative view of peer preferences, following a long-standing tradition of the market design literature (Roth, 2002; Abdulkadiroğlu and Sönmez, 2003).[4]

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 the distribution of student "abilities" at each program. We allow these preferences to be arbitrary, encompassing cases in which, for example, students wish to attend programs that: enroll the highest-ability peers, the lowest-ability peers, or peers of similar ability. Our analysis extends in a straightforward way to student preferences over the distribution of other peer characteristics.

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. As each student is "small" in our model, we show that a stable matching exists under a mild continuity condition: a sufficiently small change in the matching changes the ordinal preferences of at most a small measure of students. Unlike in Azevedo and Leshno (2016), the set of stable matchings is not generally a singleton.

Creating a stable mechanism is difficult, as it requires soliciting student preferences as functions of the sets of students attending each program.[5] Canonical, static mechanisms–such as the celebrated deferred acceptance mechanism of Gale and Shapley (1962)–in which students are only able to list ordinal

[3]Respectively these properties mean that: that no student is matched to a program dispreferred to her outside option, no seats are left unfilled at a program if some students prefer it to their assigned program, and no student is denied a seat at a program if a student with a lower program-specific score is admitted.

[4]Stability is also a desirable property from a market efficiency standpoint, as it may lead to lower attrition rates. We discuss this point later in the paper.

[5]The findings of Budish and Kessler (2021) suggest that students may not be capable of accurately stating functional preferences, and Carroll (2018) suggests that any such mechanism may be outside the realm of consideration for many centralized clearinghouses.

preferences over programs, and not over peers, will yield a stable matching in equilibrium if students have correct beliefs about the distribution of types of their peers. Students are able to "roll in" peer preferences into their reports, ensuring a stable matching. However, when students do not have accurate beliefs about the preferences of other students, these mechanisms likely fail to generate a stable matching. Our analysis shows that the presence of peer preferences makes reporting strategies dependent on distributional assumptions, thus, roughly speaking, increasing their sensitivity to incorrect beliefs.

In markets using standard mechanisms, an important consideration, therefore, is where student beliefs come from and how accurate they are. We focus on the evolution of beliefs in a discrete-time dynamic process that mirrors status quo assignment dynamics in our empirical setting. In this "status quo" procedure, students observe the distribution of student abilities at each program in the previous cohort and then submit a (ordinal) *Rank Order List (ROL)* of programs to a centralized matchmaker. The centralized matchmaker then delivers a stable matching with respect to the ROLs. Providing information on the previous cohort's matching as a guide for current applicants is common in both decentralized and centralized higher education markets; *U.S. News and World Report* annually publishes standardized test scores of entering classes from the previous year at U.S. universities, and "fuzzy" cutoffs for admission based on standardized test scores from the previous year are published in China's higher education market (Qiu and Zhao, 2007).

Under the assumption that market fundamentals do not change over time, and that students' beliefs of the distribution of peers in the current period mirror the observed matching in the previous period, this market forms a discrete-time process similar to a tâtonnement process in exchange economies, where the distribution of student abilities serves the role of "prices." Students therefore best respond to the previous period's "prices," as in the Cournot updating procedure.[6] We refer to this status quo procedure 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 is (approximately) in steady state if and only if the market creates a (approximately) stable matching.

We identify three shortcomings of the TIM process. First, it need not converge, meaning it will never generate a stable matching. We show theoretically that this process can cycle as in Scarf (1960), wherein a program enrolls a higher ability set of students in some years, and a lower ability set of students in others. This outcome is not pathological; the cyclic pattern has been observed empirically in markets operating under the TIM process. [7] Moreover, we show in a formal sense that even detailed knowledge

[6]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.

[7]For example, "great and small years," or cycles in which universities alternate between having more and less competitive student bodies, have been noted in China's higher admissions market. Specifically, students are told in an application guide, "if the university has a history of great and small years, you should pay particular attention to this cyclic factor" when submitting ROLs (p. 210 Qiu and Zhao, 2007). We are indebted to Yan Chen for this reference, and for the translation from Mandarin.

of the "functional form" of peer preferences is unlikely to be enough to prognosticate whether the TIM process will converge, and cycles are possible in most markets. Second, even if the TIM process does converge, it need not do so immediately, and therefore, instability persists along the path to stability. Third, the TIM process is fragile to changes in the market; if, for example, programs enter or exit over time, then little information may be transmitted across cohorts.

We propose a mechanism that more induces a tâtonnement process *within* each cohort of students, and does not require detailed information about the "functional form" of peer preferences. We show that this mechanism improves upon the status quo TIM procedure in all three ways, and also has desirable incentive properties, suggesting that it may not disadvantage unsophisticated students.

Other papers have studied peer preferences in a centralized matching framework. One literature focuses on the effects of couples in matching markets (see e.g. Roth and Peranson, 1999; Klaus and Klijn, 2005; Kojima, Pathak, and Roth, 2013; Nguyen and Vohra, 2018). These papers differ from ours in that peer preferences depend only on the presence of an agent's spouse, not on the entire set of peers. Another literature (Echenique and Yenmez, 2007; Bykhovskaya, 2020; Pycia, 2012; Pycia and Yenmez, 2019) studies general forms of peer preferences with small sets of students, and they primarily focus on identifying conditions under which stable matchings exist. Unlike in our setting, stable matchings frequently do not exist. Recent research also studies stability with peer preferences in large, finite, matching markets (Greinecker and Kah, 2021), and our model is most similar to that in a contemporaneous paper (Leshno, 2021), which also studies a continuum market.

There are several key differences between our paper and Leshno (2021). First, our model allows students to have preferences over the entire distribution of peer abilities, whereas Leshno's assumes students care only about summary statistics of student abilities. We show that certain reasonable forms of peer preferences cannot be expressed via (any finite number of) summary statistics. Second, we adapt our model to the case in which students have preferences only over summary statistics to provide testable results for our empirical setting. Importantly, our construction of summary statistics differs from that of Leshno; we provide results when students have preferences over their ordinal rank in the class, which reflects our empirical setting but are not supported in Leshno's analysis. Third, other than initial existence results, the focuses of our papers diverge. Leshno shows that the continuum model is a valid approximation of large, finite models while we study the consequences of peer preferences in present-day school choice markets.

We empirically investigate the presence and impact of peer preferences on stability in centralized market for college admissions in NSW, Australia's largest state. NSW matches students to programs using the (student-proposing) deferred acceptance algorithm. Students are ranked predominantly based on the *Australian Tertiary Admissions Rank (ATAR)*, which is primarily based on standardized testing. Each student submits a ROL over programs. As in our model, students receive information on the

ATAR scores–and hence ability distribution–of the cohort admitted to each program in the *previous* year. We refer to this going forward as the *Previous Year's Statistic (PYS)*. Our dataset includes the universe of applicant ATAR scores, applicants' ROLs, and program PYSs in NSW from 2003 to 2016.

We find that students' ROLs respond systematically to information about prospective peer ability, and information about their own ability. We identify that students value programs enrolling higher ability peers on average, but that this is tempered by concerns over their *relative* ability in the class (as in Frank, 1985; Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012; Tincani, 2018). The peer utility effect is asymmetric, and similar to the analogous function in Card et al. (2012); students face a utility loss only if their ATAR 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. The effect of these preferences is large: we estimate that 25% of students would be matched to a different program if we "removed" the peer component of preferences.

We use the data to formally argue for the existence of big-fish peer preferences in two ways. First, we look across time at the response of students to changes in programs' PYSs. In line with big-fish preferences, as a program's PYS increases, it receives fewer applications from students with lower ATAR scores. Moreover, the response is asymmetric; students with scores greater than the PYS are no less likely to rank the program.

Second, we look *within* the same student over time. An important feature of this market is that students submit an ROL before learning their own ATAR score, and can modify the ROL after receiving their score.[8] Big-fish preferences predict that a student will adjust her ROL to prioritize programs with similar PYSs after learning her true ATAR score. We observe that after learning their ATAR scores, students systematically *drop* programs with PYSs far above their own ATAR score, *add* new programs with PYSs closer to their own ATAR score, and *switch* the rankings of programs previously on their ROL by *promoting* programs that have PYSs closer to their own ATAR score. This switching behavior in particular is difficult to explain by consideration of admissions probability, given the strategy-proofness of the matching mechanism.

We investigate, but find little evidence to support, the following alternative models for the "big-fish" behavior we observe: students incorrectly believe that they are penalized if they are rejected from programs they rank near the top of their ROL, students optimally gather costly information over programs with ATAR scores closer to their own and prefer these programs due to greater knowledge of their underlying preferences, other non-standard preferences (such as loss aversion) studied in the literature, and students use the ATAR as a signal of the "mismatch" between their own type and that of the program. We cannot distinguish between "direct" preferences over peers, or "indirect" factors that are frequently

[8]Nei and Pakzad-Hurson (2021) also discuss how learning new information that affects preferences can impact the stability of a higher education market.

unmodeled in matching papers, such as career concerns or uncertainty about future financial opportunities.[9] However, our results apply to any setting in which the inclusion of the peer ability distribution into students' utility functions captures students' ordinal preferences at the time of application.

Although we are the first to our knowledge to show the existence of peer preferences at the university level, recent papers find that peer preferences exist at the high school level (Rothstein, 2006; Beuermann et al., 2019; Allende, 2020), and matter above and beyond value-added measures (Abdulkadiroğlu et al., 2020; Beuermann and Jackson, 2019). These papers find that (parents of) students prefer, on average, programs where peers have higher ability. Our analysis corroborates this finding, but importantly differs in that we study how a student's *relative* ability affects the desirability of a program. Previous papers in this literature generally do not consider relative peer preferences as their models assume a constant effect of peer scores on the preferences of all students.[10] That our data allow us to observe student ROLs before and after learning their relative abilities is unique in this literature, and our analysis reveals a more nuanced "functional form" of peer preferences in the NSW market than has been presented in the existing literature.

The "functional form" and root causes of peer preferences potentially depend on a number of factors–such as cultural norms[11] and student autonomy[12]–that differ across markets (for a similar thesis, see Sacerdote, 2014). Our main message is agnostic to the form or cause of peer preferences; the test we derive for stability in the presence of peer preferences relies on studying the convergence (or lack thereof) of within-program student ability over time, and this test applies to a general class of functional forms of peer preferences.

We empirically study the long-run stability of the NSW 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 in 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 the 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

[9] Dillon and Smith (2017) find that students with more information about financial aid opportunities are more likely to select programs with higher-ability peers.

[10] Beuermann et al. (2019) and Abdulkadiroğlu et al. (2020) separately estimate preferences for average peer ability separately for high- and low-ability students, which is closer to an analysis of relative peer preferences. They find more muted effects of high peer ability on program desirability for low ability students, which is consistent with big-fish preferences.

[11] A long literature establishes the so-called "tall poppy" syndrome in Australia, wherein students react negatively to those who overachieve relative to peers (see, e.g. Feather, 1989), possibly leading Australians to avoid programs they perceive their peers to be "overachieving."

[12] Pop-Eleches and Urquiola (2013) and Ainsworth et al. (2020) study the same high school admissions market and find that while students at the bottom of the ability distribution at a program suffer from psychic costs, but that parents prefer to send their children to programs with higher-achieving students. Teske, Fitzpatrick, and Kaplan (2007) find that older students have more involvement in school choice decisions; it stands to reason that in college admissions markets in which students have more autonomy, the direction of peer preferences could differ.

there is long-run stability at the "top" of the market, but not at the bottom. This instability is associated with higher attrition rates, and particularly impacts low-socioeconomic status students.

Second, we follow Epple and Romano (1998) and Avery and Pathak (2021) and study a market in which students prefer programs with higher ability peers. Unlike the NSW market, we show that a stable matching is generated in every period–with or without entry and exit of programs–when student preferences do not depend on other characteristics of schools.

The remainder of the paper is structured as follows: Section II introduces our model and main theoretical results, and discusses the long-run stability of two markets; Section III discusses details of the NSW Tertiary Education System and provides evidence of peer preferences; Section IV empirically analyzes the (lack of) stability present in the NSW market; Section V concludes. Omitted proofs and additional results are relegated to the Appendix.

II Model

II.A Setup

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

To capture that student preferences depend on their peers, we characterize potential peer groups (that need not respect feasibility) as a useful building block. An *assignment* of students to programs 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 is assigned to a subset of programs, Condition 2 states that a program is assigned to a 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. We denote the set of all assignments by $\mathcal{A}$.

Each student is characterized by $\theta=(u^\theta,r^\theta)$. $u^\theta(c|\alpha)\in\mathbb{R}$ represents the cardinal utility the student derives from being assigned to only program c given that other students are assigned according to assignment $\alpha\in\mathcal{A}$. That is, $u^\theta(c|\alpha)=u^\theta(c|\alpha(\theta)=c \text{ and } \{\alpha(\theta')\}_{\theta'\in\Theta\setminus\{\theta\}})$. We normalize $u^\theta(c_0|\alpha)=0$ for all $\theta\in\Theta$ and all $\alpha\in\mathcal{A}$, 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 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 $\eta\in H(N)$, $\eta\{\theta|r^{\theta,c}<y\}=y$ for all $y\in[0,1]$

and all $c \in C$, that is, the marginal distribution of every program's rankings is uniform. Therefore, the set of all student types is $\Theta = \mathbb{R}^{N+1} \times \mathcal{A} \times [0,1]^{N+1}$.

We denote a market by $E = [\eta, q, N]$.

It will often be useful to denote the ordinal preferences of student 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 $\lambda^{c,x}(\alpha) := \eta(\{\theta | r^\theta \leq x \text{ and } \theta \in \alpha(c)\})$. 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. 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 $B_\epsilon(r)$ be the set of points within ϵ distance of $r \in R$, $B_\epsilon(r) := \{r' \in [0,1]^{N+1} | \epsilon > ||r - r'||_\infty\}$. Then there exists $\omega > 0$ such that for any $\alpha \in \mathcal{A}$, any $c \in C \setminus \{c_0\}$, and any $r \in R$, $\eta(\{\theta \in \Theta | r^\theta \in R \cap B_\epsilon(r) \text{ and } c \succ^{\theta|\alpha} c' \text{ for all } c' \in C \setminus \{c\}\}) > \omega \eta(\{\theta \in \Theta | r^\theta \in R \cap B_\epsilon(r)\})$.

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

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

A4 Peer preferences are *aggregate unresponsive*: 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 $\eta(\{\theta \in \Theta | \succeq^{\theta|\alpha} \neq \succeq^{\theta|\alpha'}\}) < \epsilon$.[13]

A2 and **A4** are richness assumptions; **A2** states that for any assignment, there exist some students across the score distribution who most prefer each program; **A4** states that the ordinal rankings of the vast majority of students do not change for small changes in the composition of peers.

[13] Leshno (2021) refers to a similar condition as *diversity of preferences*.

II.B Stable matchings

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. To reduce a multitude of essentially identical matchings that differ only for a measure zero set of students, throughout the paper 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 *(pairwise) stable* if there do not exist any student-program blocking pairs.[14]

We specify a class of assignments defined by admission cutoffs. This construction will be used to characterize stable matchings, as in Azevedo and Leshno (2016). A cutoff vector $p \in [0,1]^{N+1}$ is subject to $p^{c_0} = 0$. One can construct an assignment given a cutoff vector p as follows. First, fix an arbitrary assignment α', and corresponding ability distribution $\lambda = \lambda(\alpha')$. Second, let each student θ choose her favorite program among those where her program-specific score is weakly above the cutoff.[15] We refer to this program as the *demand of* θ, and denote it by

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

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

We similarly define the *demand for program* c as

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

[14] Throughout, we shorten the name of this solution concept to "stability." The following axioms are jointly equivalent to stability: A matching μ is: *individually rational* if $\mu(\theta) \succeq^{\theta|\mu} c_0$ for all θ; *non-wasteful* if for some θ and c it is the case that $c \succ^{\theta|\mu} \mu(\theta)$ then $\eta(\mu(c)) = q^c$; *fair* if there does not exist θ, θ' and c such that $\mu(\theta') = c$, $c \succ^{\theta|\mu} \mu(\theta)$, and $r^{\theta,c} > r^{\theta',c}$. Our analysis is largely unchanged other that notational complications by relaxing non-wastefulness to allow a program to reject sufficiently low-scoring students even when it has an excess supply of seats.

[15] 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 $D^{\theta}(p, \lambda) \in \underset{\succeq^{\theta|\lambda}}{\operatorname{argmax}} \{c \in C | r^{\theta,c} \geq p^c\}$. As this is relevant for at most a measure zero set of students, for notational simplicity we proceed as if each student has a unique top choice.

The assignment $\alpha = A(p,\lambda)$ is defined by setting $\alpha(\theta) = D^{\theta}(p,\lambda)$ for every student θ. By construction, each student 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 $c \in C$*

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

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 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) satisfies* 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 satisfies 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 $(\hat{p},\lambda)$ is market clearing and satisfies 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 markets. We construct an operator whose fixed point corresponds to a stable matching, and show, using a fixed-point theorem, that this repeated iteration of this operator yields a stable matching.[16]

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

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

[16]Our proof generalizes a fixed-point argument of Leshno (2021). Grigoryan (2021) uses the same fixed-point theorem in a matching market with complementarities. More broadly, fixed-point arguments are often used to show existence results in the literature (see, for example, Pycia and Yenmez, 2019).

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

We show this result via Example 3 in the appendix. 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.

II.C Canonical Mechanisms and Stability

Theorem 1 tells us that a stable matching exists in many markets. Can a market designer ensure one using a "canonical" matching mechanism? We show that the answer is "no" in many settings.

In any market E, define a *one-shot matching mechanism* φ as a simultaneous-move, deterministic game in which each student θ submits a ROL $\succ^\theta$ over programs $c \in C$. φ maps ROLs $\succ = \{\succ^\theta\}_{\theta\in\Theta}$ and scores into a matching, that is $\varphi : (\mathcal{P} \times [0,1]^{N+1})^\Theta \to \mathcal{M}$. We represent the resulting matching from report $\succ$ as $\varphi(\succ)$, the matched partner for student θ as $\varphi^\theta(\succ)$, and the set of students matched to program c as $\varphi^c(\succ)$. A one-shot mechanism φ *respects rankings* if for any $\succ$ the following is satisfied: if $r^{\theta,c} > r^{\theta',c}$ for all c and $|\{c|c\succ^\theta \varphi^{\theta'}(\succ)\}| \leq |\{c|c\succ^{\theta'} \varphi^{\theta'}(\succ)\}|$, then $\varphi^\theta(\succ)\succeq^\theta \varphi^{\theta'}(\succ)$. That is, a student is not matched to a program she ranks below program c if she has a higher score (across all programs) than another student who is matched to c and she ranks c at least as high as the student with lower scores.[17] Informally, we refer to φ as "canonical" if it is a one-shot mechanism which respects rankings. A stronger requirement is stability. A one-shot mechanism φ is *stable* if for any $\succ$, $\varphi(\succ)$ is stable *with respect to* $\succ$. Note that any one-shot stable mechanism φ must respect rankings.[18]

The following result says that we can expect a clearinghouse to generate a stable matching by using a stable mechanism if students have full knowledge of the distribution of student types.[19] In this case, the set of stable matchings is Bayes Nash implemented by any stable mechanism φ as students are able to "roll in" peer considerations into their ROLs. That is, for any stable matching μ_*, there is an equilibrium in which each student θ reports $\succ^\theta = \succeq^{\theta|\mu_*}$.[20] On the other hand, if students' beliefs about the distribution of types is sufficiently misspecified, then we should not expect a clearinghouse

[17] The requirement that she ranks c at least as high as the student with lower scores (i.e. $|\{c|c\succ^\theta \varphi^{\theta'}(\succ)\}| \leq |\{c|c\succ^{\theta'} \varphi^{\theta'}(\succ)\}|$) is included to expand the class of covered mechanisms to include the immediate acceptance mechanism. Removing this additional requirement would not otherwise change our results.

[18] Proof: Suppose not. Then for some $\succ$ there exist θ, θ' with $r^{\theta,c} > r^{\theta',c}$ for all c, and $c^* = \varphi^{\theta'}(\succ)\succ^\theta \varphi^\theta(\succ)$. But since $r^{\theta,c^*} > r^{\theta',c^*}$, it is the case that (θ,c^*) form a blocking pair. Contradiction with φ being stable.

[19] Full knowledge of the distribution of types is not a necessary condition for the clearinghouse to generate a stable matching. As the distribution of peers within programs is the only payoff relevant feature of the market (Esponda and Pouzo, 2016) (in a strategy-proof mechanism), a stable matching can be generated in equilibrium if students anticipate the distribution of peers at each program with sufficient accuracy. We explore this in Section II.D.

[20] As we discuss in the proof of Proposition 1, for any stable mechanism φ, if almost all students θ report $\succ^\theta = \succeq^{\theta|\mu_*}$ then $\varphi(\succ) = \mu_*$, as μ_* is the only stable matching associated with these preferences. Moreover, we show the existence of an equilibrium yielding μ_* in which each student lists only one program as acceptable. Therefore, 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 in equilibrium, under full knowledge of the distribution of student types.

to generate a stable matching using a canonical mechanism.

Suppose student θ believes the measure over student types is given by $\sigma^\theta \in H(N)$. Let $\succsim$ be a strategy profile, and let $\mu(\sigma^\theta,\succsim)$ be the anticipated matching of student θ. Then θ's expected ordinal rankings over programs given σ^θ and $\succsim$ is $\succeq^{\theta|\mu(\sigma^\theta,\succsim)}$. We say that student θ *lacks rationality for the top choice at* $(\succsim,\varphi)$ if the $\succeq^{\theta|\mu(\sigma^\theta,\succsim)}$-maximal program is not a $\succeq^{\theta|\varphi(\succsim)}$-maximal program. For any $r \in [0,1)^{N+1}$ let $L_{\succsim,\varphi,r} := \{\theta | r^\theta \geq r$ and θ lacks rationality for the top choice at $(\succsim,\varphi)\}$.

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

1. *Let φ be stable and suppose $\sigma^\theta = \eta$ for all $\theta \in \Theta$. Then the set of all stable matchings of market E is identical to the set of all Bayes Nash equilibrium outcomes of φ.*

2. *Let φ respect rankings and let μ_* be a stable matching. If for all $r \in [0,1)^{N+1}$ and all $\succsim$ it is the case that $\eta(L_{\succsim,\varphi,r}) > 0$ 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; a number of additional conditions must be met. First, these students must have sufficiently strong peer preferences so that their incorrect beliefs change their ROLs. Second, these students must have scores above the admission thresholds at these programs. Third, the incorrect beliefs affect 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 across the ability distribution.

II.D Tâtonnement with Intermediate Matching and Belief Updating

Given Proposition 1, an important question is how students form beliefs when submitting ROLs to a centralized mechanism. We model belief formation in a tâtonnement-like process, in which beliefs update given the assignment of the previous cohort of students.

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 $t,t' \geq 1$, markets E_t and $E_{t'}$ are identical, that is, the measure over student types and program capacities are constant over time. We therefore omit all time indices when denoting market fundamentals.

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 as follows.

The market is initialized with an arbitrary assignment $\mu_0 \in \mathcal{A}$. We initialize the market with an assignment instead of a matching so as not to require students in the first cohort to be fully informed of all particulars in the market, for example, the capacity at each program; our results are qualitatively unchanged if we instead allow students to have (potentially heterogeneous) beliefs over the initial

assignment μ_0, but the exposition would become more cumbersome with this additional generality. Incoming students at time t observe μ_{t-1}. A matchmaker solicits an ROL from each student, and then uses a stable matching mechanism to construct matching μ_t.

An instructive observation moving forward is that there is a unique stable matching μ_t in a market in which preferences are defined by μ_{t-1}. This result follows directly from Assumption **A2** and Theorem 1 of Azevedo and Leshno (2016). 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 given assignment α.

Remark 2. *Fix a market $E = [\eta, q, N]$ and let μ_t be the matching constructed at time $t \geq 1$ in the TIM process. Then μ_t is the unique stable matching in market $E' = [\zeta^{\eta,\mu_{t-1}}, q, N]$.*

By this remark, μ_t is the outcome of any stable matching mechanism at time t in the TIM process. One such mechanism is student-proposing deferred acceptance, which is also strategy-proof in the static setting.[21] We assume (and later show empirical evidence that) students use information from the previous period in a Cournot-updating fashion, that is, each period t student θ assumes that μ_t will equal μ_{t-1}. Therefore, each student θ has a weakly dominant strategy to submit her "true" preferences $\succeq^{\theta|\mu_{t-1}}$ in *any* stable matching mechanism, and we adopt the assumption that students report their true preferences going forward.

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 $\lambda_t = \lambda(A(p_t, \lambda_{t-1}))$. 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, $S(p, \lambda') = \lambda(A(p, \lambda'))$.

Let $\lambda_0 := \lambda(\mu_0)$. For any $t \geq 1$, $\mu_t = A(p_t, \lambda_{t-1})$, where $p_t = P\lambda_{t-1}$ and $\lambda_{t-1} = S(p_{t-1}, \lambda_{t-2})$. If $\mu_t = \mu_{t+1}$ then 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 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

[21] See Abdulkadiroğlu, Che, and Yasuda (2015) for further details on this mechanism in the continuum model.

in "approximate" steady state does the TIM process generate an "approximately" stable matching. Before stating the result, we state a definition of ϵ-stability, which requires that fewer than ϵ share of students are involved in a blocking pair. Our notion of approximate stability comes from selecting a small ϵ.

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, $\eta(\{\theta|(\theta,c) \text{ block } \mu \text{ for some } c \in C\}) < \epsilon$.*

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

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.*

Consider an observer who does not necessarily know the preferences students have over peers and who 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 process necessarily converge?

A natural question arises: (if market E has a stable matching) does the TIM process always converge for any μ_0? If so, then the TIM process guarantees an (approximate) stable matching in the long run, and a patient market designer may be content to rely on this process.

We show that even a very detailed understanding of the functional form of peer preferences is unlikely sufficient to accurately forecast whether the market will converge to stability. We study the class of peer preferences which are separable across programs, that is, students care only about the distribution of student types at their own program.[22] We find for any collection of such peer preferences, there exist markets for which the TIM process converges. We further find that within this separable class, any collection of peer preferences that admits a *negative externality group*–informally, a set of students who reduce one another's utilities–admit markets in which the TIM process does not converge.

Focusing on peer preferences that admit a negative externality group is important for two reasons. First, we show that the existence of a negative externality group is likely from a topological perspective; separable peer preferences generically admit a negative externality group. Second, negative externality

[22]This rules out situations in which students care about students at other programs. Sasaki and Toda (1996) study such a problem in a one-to-one matching framework.

groups are economically meaningful, as they capture the notion that there can exist certain students who are undesirable to others.

We formalize the machinery necessary to state this result. Let $\hat{H}(N)$ be the set of measures over student types with N programs such that for almost all students θ and all programs $c \in C \setminus \{c_0\}$ we can represent $u^\theta(c|\alpha) = v^{\theta,c} + f^{\theta,c}(\lambda^c(\alpha))$, where $v^{\theta,c}$ is an "intrinsic" component of preferences, and $f^{\theta,c}(\cdot)$ is uniformly continuous and uniformly bounded: for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $\theta \in \Theta$ and any $c \in C \setminus \{c_0\}$, if $\lambda^c, \hat{\lambda}^c \in \Lambda$ satisfy $||\lambda^c - \hat{\lambda}^c||_\infty < \delta$ then $|f^{\theta,c}(\lambda^c) - f^{\theta,c}(\hat{\lambda}^c)| < \epsilon$, and there exists $a < b$ such that $f^{\theta,c}(\cdot) \in [a,b]$ for all $\theta \in \Theta$ and $c \in C$. We write $f \mapsto \eta$ if the preferences of almost every student θ induced by η can be represented by a uniformly continuous and uniformly bounded family of functions $\{f^{\theta,c}(\cdot)\}_{\theta \in \Theta, c \in C \setminus \{c_0\}}$.

We first show that if $\eta \in \hat{H}(N)$ and the TIM process does not converge, it is not due to a lack of existence of a stable matching.

Remark 3. *There exists a stable matching in any market $E = [\eta, q, N]$ such that $\eta \in \hat{H}(N)$.*

We say that $\eta \in \hat{H}(N)$ does not admit a *negative externality group* if there does not exist any $c \in C \setminus \{c_0\}$, any $\alpha(c)$, and any positive measure sets of students $\Theta^I \subset \alpha(c)$ and $\Theta^O \subset \Theta \setminus \alpha(c)$ with $\eta(\Theta^I) \geq \eta(\Theta^O)$ such that $f^{\theta,c}(\lambda^c(\Theta^O \cup \alpha(c) \setminus \Theta^I)) > f^{\theta,c}(\lambda^c(\alpha(c))$ for all $\theta \in \Theta^I$. In words, no set of students Θ^I can prefer a program c when a smaller set of students Θ^O replaces them. Note that $\eta(\Theta^I)$ can be arbitrarily small, or even equal to zero, implying that this condition can be satisfied if students in Θ^I prefer that some of their peers do not attend program c without replacement.

To establish what it means for two collections of peer preferences to be "close," we first normalize the functional forms of peer preferences to address issues of multiplicity.[23] We say that $\{F^{\theta,c}(\cdot)\}_{\theta \in \Theta, c \in C \setminus \{c_0\}}$ is a *canonical* representation of η if $F \mapsto \eta$ and for almost every θ and every $c \in C \setminus \{c_0\}$ there exists an assignment α such that $F^{\theta,c}(\lambda(\alpha)) = a$ and if there exists assignment α' such that $u^\theta(c|\alpha') > u^\theta(c|\alpha)$ then there exists an assignment α'' such that $F^{\theta,c}(\lambda(\alpha'')) = b$. That is, F is a canonical representation if it renormalizes the peer preference component of utilities such that for almost all student-program combinations, the student's least favorite assignment at that program yields the smallest allowable value of F and the student's favorite assignment at that program (assuming the student is not indifferent between all assignments) yields the largest allowable value of F.

For two measures $\eta, \tilde{\eta} \in \hat{H}(N)$ and respective canonical representations, let

$$||\eta - \tilde{\eta}||_p := \inf_{F \mapsto \eta, \tilde{F} \mapsto \tilde{\eta}} ||F^{\theta,c}(\lambda(\cdot)) - \tilde{F}^{\theta,c}(\lambda(\cdot))||_\infty.$$

The "$||\cdot||_p$" norm calculates the pointwise supremum difference between canonical representations

[23] That is, there are renormalizations of the function $u^\theta(c|\alpha) = v^{\theta,c} + f^{\theta,c}(\lambda^c(\alpha))$ that yield the same ordinal preferences for student θ over programs for each α.

$\{F^{\theta,c}(\cdot)\}_{\theta\in\Theta,c\in C\setminus\{c_0\}}$ and $\{\tilde{F}^{\theta,c}(\cdot)\}_{\theta\in\Theta,c\in C\setminus\{c_0\}}$. The "inf" term selects canonical representations that are the "closest" two canonical representations of η and $\tilde{\eta}$, respectively, which by our construction, removes from consideration canonical representations that differ for only a zero measure subset of students.

Theorem 3.

1. *For any $N\geq 1$ and any $\eta\in\hat{H}(N)$ such that $f\mapsto\eta$, there exists a market $E=[\hat{\eta},q,N]$ such that $f\mapsto\hat{\eta}$ and the TIM process converges in market E for any starting condition μ_0.*

2. *For any $N\geq 1$, the set of measures that admit a negative externality group is open and dense in $\hat{H}(N)$ with respect to the $||\cdot||_p$ norm.*

3. *For any $N\geq 1$ let $\eta\in\hat{H}(N)$ and suppose that $f\mapsto\eta$. If η admits a negative externality group then there exists a market $E=[\hat{\eta},q,N]$ such that $f\mapsto\hat{\eta}$ and a starting condition μ_0 for which the TIM process does not converge in market E.*

We interpret this result through the lens of a researcher interested in the long run stability of a market utilizing the TIM procedure. Even a very detailed understanding of the functional form of peer preferences is unlikely to be enough to know whether the market will converge to stability; point 3 shows that whenever a negative externality group exists, the TIM process can cycle as in Scarf (1960). However, even the existence of a negative externality group does not imply that it is impossible for the TIM process to converge (point 1).

We further investigate when the TIM process is guaranteed to converge in two ways. First, in Appendix C we study peer preferences that are monotonically increasing in the measure of students matched to a program. We show that these preferences do not admit a negative externality group, and in any market with $N<3$ these preferences guarantee convergence of the TIM process for any starting μ_0. However, for $N\geq 3$ we show again that for any family of functions $\{f^{\theta,c}(\cdot)\}_{\theta\in\Theta,c\in C\setminus\{c_0\}}$ which are strictly increasing in the measure of students matched to a program, one can again construct a market and a starting condition μ_0 for which the TIM process will not converge.

Second, we discuss when convergence of the the TIM process can be guaranteed if further factors in the market are pinned down. In Appendix A we provide intuition for how peer preference functions, student scores, and capacities can interact to lead to convergence via two examples.

In the first example, we show that the TIM process does not always converge (even though there is a unique stable matching in the market). Understanding this lack of convergence is gained through a comparison to convergence of tâtonnement processes in exchange economies. This example exhibits an analogue of the *individual* gross substitutes condition: students disprefer having lower ability than the mean ability of their peers, and therefore, the mean ability plays the role of the price in exchange economies. As the "price" of a program rises, each student's utility from attending it decreases. However,

we have a failure of *aggregate* gross substitutes; due to capacity constraints and student unit demand, as the "price" rises, low scoring students are now able to attend the program, taking over seats from higher scoring students with weaker intrinsic preferences who decline to enroll due to the high "price." The increased demand of these low scoring students drives down the "price," thus leading to a cycle.

In a second example, we study a nearly identical market that differs only in that students have preferences over the *median* ability of their peers. Importantly, the median is not affected by outliers: given that top-ranked students enroll in the same program for each $t \geq 1$, the median is guaranteed to stay the same in the TIM process. In contrast, the mean is sensitive to the entire distribution of enrolling students: even if the same top-ranked students enroll, the mean can decrease if students with middle rankings do not enroll and some with lower scores do. Therefore, the example with the median captures the notion of aggregate gross substitutes.

As seen in these examples, convergence potentially depends on a number of factors, including program rankings over students, the functional form of peer preferences, and program capacities. We discuss sufficient conditions, which reflect two real-world markets, which guarantee the convergence of the TIM process in the following section and in Appendix E.2.

II.E New South Wales Markets

In this section we describe key restrictions that characterize the NSW market, and study the convergence properties of the TIM process in this market. We present empirical evidence to justify these restrictions in Section III.

An important consideration in NSW markets is that students have access only to a summary statistic of the distribution of student types in previous cohorts. We therefore briefly provide general theoretical results, mirroring those in the previous section, in markets in which 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 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 $\epsilon > 0$ there exists some $\delta > 0$ such that for any two assignments α, α' that satisfy $||s(\lambda(\alpha)) - s(\lambda(\alpha'))||_\infty < \delta$, we have that $\eta(\{\theta | \succeq^{\theta|\alpha} \neq \succeq^{\theta|\alpha'}\}) < \epsilon$.

A7 Fix $\epsilon > 0$. There exists $\delta > 0$ such that for any assignments α, α' that satisfy $\alpha = A(p,\lambda)$, $\alpha' = A(p',\lambda')$ for some $(p,\lambda),(p',\lambda') \in [0,1]^{N+1} \times \Lambda^{N+1}$ and $||\lambda(\alpha) - \lambda(\alpha')||_\infty < \delta$, we have that $||s(\lambda(\alpha)) - s(\lambda(\alpha'))||_\infty < \epsilon$.

Limiting attention to markets where peer preferences depend on summary statistics is not without loss of generality; in Appendix E we show that some reasonable functional forms of peer preferences cannot be represented by (any finite number of) summary statistics.

We proceed to show existence of a stable matching in markets where preferences depend on summary statistics, as in Leshno (2021). First, we note that our construction of summary statistics contrasts from that of Leshno (2021) as it enables us to study certain important functional forms of peer preferences, including "ordinal" peer preferences that we will shortly present in the characterization of NSW markets. Our definition allows summary statistics to depend on the entire ability distribution induced by an assignment. Importantly, **A7** imposes an analogue of our aggregate unresponsiveness assumption (**A4**) *only for market assignments that appear in the TIM process for* $t \geq 1$. We formalize how ordinal summary statistics satisfy this analogue of aggregate unresponsiveness in our upcoming proof of Remark 4. As can be seen in the proof of Theorem 1, our fixed-point operators only generate market assignments, yielding the following result.

Corollary 1. *Let E be a market satisfying **A1**,**A2**,**A5**$-$**A7**. Then E has at least one stable matching.*

The following result mirrors Theorem 2. In a market 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 a market satisfying **A1**,**A2**, and **A5**, and let $\mu_1, \mu_2, \ldots$ 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. *If E satisfies **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.E.1 NSW Characterization

There are two important sets of stylized facts, discussed empirically in Section III, that our modeling of the NSW 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 allow for 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. Our modeling below reflects these restrictions, although it is unlikely that they exactly and fully represent all aspects of the NSW market. However, they provide a tractable framework in which we can derive predictions on the stability of the market in the TIM process.

As before, let a market be characterized by $E=[\eta,q,N]$ where $\eta \in H(N)$ 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 $q^{c_0} \geq 1$. Let $E_1,E_2,...$ be a sequence of markets, where for each $t \geq 1$ there is a set $C_t \subset C$ of active programs, where $|C_t|=N_t+1$ and $c_0 \in C_t$. $E_t := [\eta,q_t,N_t]$ 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 each market E_t satisfies **A1**. We formalize the stylized restrictions on preferences with the following three points:

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

AA2 Big-fish preferences: Each student θ has utility function $u^\theta(c|\alpha) = v^{\theta,c} - f^{\theta,c}(r^\theta, s^c(\lambda(\alpha)))$, where, for all $c \in C_t \setminus \{c_0\}$, $f^{\theta,c}(\cdot,\cdot) \geq 0$, 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: $s^c(\cdot)$ is said to represent the $(k^c)^{th}$ *highest score* if there exists $k^c \in [0,1]$ such that for any market clearing matching $\mu \in M_t$, $t \geq 1$, with associated cutoff vector $p \in [0,1]^{N_t+1}$, and any α with $\lambda^c(\alpha)=\lambda^c(\mu)$, $s^c(\lambda(\alpha))$ equals the supremum value $z \in [0,1]$ for which $\eta(\{\theta' \in \alpha(c) | r^{\theta'} > z\}) = k^c$ (if such a number exists, and p^c otherwise). For each $t \geq 1$ and each $c \in C_t$ there exists $k^c > 0$ where $s^c(\cdot)$ represents $(k^c)^{th}$ highest score.

AA1 reflects the fact that a standardized score determines admission to programs. **AA2** states that students face an additive peer cost when assigned to a program in which 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 assignments. 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 in our upcoming results for certain markets.[24]

[24]Note that the TIM process only produces matchings $\mu \in M_t$. Therefore the sequence of matchings generated from two otherwise identical markets will be identical if their summary statistic vectors coincide on this restricted set of matchings. This means that a wider class of summary statistics falls into the category of the 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 the k^{th} highest

The following reflects our stylized facts regarding 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 $B_1 \cup B_2 = C \setminus \{c_0\}$. We place no restrictions on the relative sizes of these two blocks. 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: $v^{\theta,c} > v^{\theta,c'}$ 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 $[0,1]$. There exists $\omega > 0$ such that for any $\alpha \in \mathcal{A}_t$ and any $c \in B_1$, $\eta(\{\theta \in \Theta | r^\theta \in R \text{ and } c \succ^{\theta|\alpha} c' \text{ for all } c' \in C_t \setminus \{c\}\}) > \omega \eta(\{\theta \in \Theta | r^\theta \in R\})$.

Certain programs are long lived (**AA4**) and these are precisely the more desirable programs (**AA5**).[25] **AA6** is a relaxation of **A2**, ensuring full support of preferences over top-block programs. Combining the new restrictions in this section leads to our definition of a NSW market.

Definition 5. *A sequence of markets $E_1, E_2, \ldots$ is said to be* NSW *if it satisfies **A1**, **AA1**-**AA6**.*

There exists a unique stable matching for each E_t, $t > 0$ in a NSW market. We provide a pseudo-serial-dictatorship mechanism in the appendix that serves as a constructive proof of existence. A student θ with a sufficiently high score receives the same partner in the stable matching for each market E_t, $t \geq 1$.

Proposition 2. *Let $E_1, E_2, \ldots$ be NSW, and consider any $t \geq 1$. 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. *for any student θ and any $c \in B_1$ such that $r^\theta \geq s^c(\lambda(\mu_t^*))$, it is the case that $\mu_t^*(\theta) = \mu_{t'}^*(\theta)$ for all $t' \geq 1$.*

We now discuss the TIM process with entry and exit, which is largely analogous to that in 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.

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 $k^c := m \cdot \eta(\mu(c))$. Therefore, summary statistics such as the median (see Example 2) 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 $c \succ^{\theta|\alpha} c_0$ 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 $c \in C_t \setminus \{c_0\}$, $\eta(\mu(c)) = q^c$.

[25] Condition **AA5** is based on block-correlated preferences, presented in Coles, Kushnir, and Niederle (2013).

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 $\lambda_{t-1} = \lambda(\nu_{t-1})$. The initial assignment $\nu_0 \in \mathcal{A}_1$ 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.

Regardless of entry and exit, the summary statistics of popular programs (those in block B_1) converge to their stable levels in the TIM process, and except in rare cases, this convergence occurs in finite time. Let $V := \{\theta | r^\theta \geq \min_{c \in B_1} s_*^c\}$ 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 students $\theta \in V$ eventually receive their stable matching partner.

Theorem 4. *Let $E_1, E_2, \ldots$ be NSW. For any $\nu_0 \in \mathcal{A}_1$ there generically exists some time $T < \infty$ such that in the TIM process, $s_t^c = s_*^c$ for all $c \in B_1$ and $\eta(\{\theta \in V | \mu_t(\theta) = \mu_*(\theta)\}) = \eta(V)$ for all $t > T$.*

In contrast to the top programs and top students, 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, long-run instability only affects students with scores below the stable matching summary statistics of all programs in the top block.

The question arises of how much instability is caused by entry and exit. We study this question in the appendix, and show that low-scoring students are the only ones directly negatively affected by instability in the long run.

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 or exit. Theorem 4 implies that the TIM process converges to the (unique) stable matching in finite time, and our empirical test of convergence (Corollary 2) will hold.

Remark 4. *Let $B_1 = C \setminus \{c_0\}$, and let $E = E_1 = \ldots$ be NSW. 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 $\inf_{\theta \in \alpha(c)} r^\theta$ otherwise). Then E also satisfies **A2**, **A5**-**A7**.*

II.F A More Stable Mechanism

A mechanism utilizing the algorithm constructed in the Proof of Theorem 1 may be infeasible as it requires soliciting student preferences as functions of the sets of students attending each program.[26] We instead present a constrained mechanism that does not rely on detailed information about the functional form of peer preferences and only requires students to submit ROLs as in the TIM process. This

[26] Budish and Kessler (2021) suggest that students may not be capable of accurately stating functional preferences, and Carroll (2018) suggests that any such mechanism may be outside the realm of consideration for many centralized clearinghouses.

mechanism does not run across years, and instead attempts to find or approximate a stable matching for each cohort of students. Unlike the TIM process, it suffers neither from 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.

Students in each cohort are assigned to one of many smaller submarkets, and students in each submarket submit ROLs sequentially after seeing the previous submarket's ability distribution. Formalizing this mechanism involves specifying the students in each submarket, 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" to refer to a generic submarket below to be evocative of the time index in the TIM process.

First, we specify students in each submarket E_t, $t \in \{1,...,T\}$. Let η_t represent the submarket t measure over Θ, where $\sum_{\ell=1}^{T} \eta_t(\Theta) = 1$. We will write $\theta \in E_t$ to mean that student θ is active in submarket E_t. Each η_t is constructed "uniformly at random," that is, for any open set $\Theta^o \subset \Theta$, it is the case that $\eta_t(\Theta^o) = \eta(\Theta^o) \cdot \eta_t(\Theta)$. We assume $\eta_t(\Theta) \to 0$ for all t as $T \to \infty$.

Second, we specify the programs. Each program $c \in C$ is active in each submarket, and has a submarket t specific capacity constraint $q_t^c = q^c \cdot \eta_t(\Theta)$ seats available. The capacity vector in submarket t is q_t. Combining these components, we denote submarket t as $E_t = [\eta_t, q_t, N]$.

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

We now formally define our proposed iterative matching mechanism.

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 students $\theta \in E_t$ the distribution $\lambda(\mu_{\tau-1})$ and solicit ROLs of programs. Using a stable mechanism, 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 students the distributions $\lambda(\mu_{\tau-1})$ and solicit their ROLs over programs. Using a stable mechanism, create matching $\mu_{(\mu_0,\delta)}$ in aggregate market E, which 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 (but does not affect the final matching generated). We denote the outcome of the TFM mechanism (assuming the mechanism terminates) for a given (μ_0, δ) as $\mu_{(\mu_0,\delta)}$.

The TFM mechanism terminates if the TIM process converges, and for sufficiently small δ, it creates a nearly-identical matching. Moreover, the TFM mechanism can create a nearly-stable matching even when the TIM process does not converge. We prove this by construction, which may be of independent interest–we show that the TIM process potentially suffers from a lack of local convergence; even if the TIM process 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 also has desirable incentive properties. We say that a student $\theta \in \Theta_t$ *misreports at step t* if she submits a preference profile $\succsim^{\theta} \neq \succeq^{\theta|\mu_{t-1}}$. 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 uniformly continuous in λ,[27] 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.[28] Because of this, the TFM mechanism potentially levels the playingfield between "sophisticated" students who submit ROLs best responding to the strategies of others, and "sincere" students who are unwilling or unable to misreport (for studies on the role of mechanisms in leveling the playingfield between sophisticated and sincere players see egs. Pathak and Sönmez, 2008; Song, Tomoeda, and Xia, 2020).

Finally, we show that for any μ_0 and δ there is sufficiently large T such that if the TFM mechanism terminates, it does so with no student being asked her preferences more than twice, and an arbitrarily large share of students being asked only once. Recalling that T does not affect the final matching generated, this implies that for large enough T there are small additional reporting costs associated with this mechanism over canonical, one-shot mechanisms.

Proposition 3. *Let E be a market satisfying **A4** and let $\epsilon > 0$.*

1. *Suppose that for a given μ_0, the TIM process converges to (stable) matching μ_*. Then for any stopping rule δ the TFM mechanism terminates in market E with starting condition μ_0, and there exists $\delta^* > 0$ such that for any stopping rule $\delta < \delta^*$, $\mu_{(\mu_0,\delta)}$ is $\epsilon-$stable.*

[27] That is, if for all $\gamma_1 > 0$ there exists $\gamma_2 > 0$ such that for (almost) all $\theta \in \Theta$ and all $c \in C$, $|u^{\theta}(c|\lambda) - u^{\theta}(c|\lambda')| < \gamma_1$ for any $\lambda, \lambda' \in \Lambda^{N+1}$ with $||\lambda - \lambda'||_{\infty} < \gamma_2$.

[28] This mechanism does admit "babbling" equilibria in which some students report arbitrary preferences in early periods because they anticipate being able to correct their reports. Note, however, that students have no strict incentive to do this, as no student's report affects any final matching, except their own. One way to remove such equilibria is to obscure the order in which students are solicited to submit preferences (assuming students do not fully coordinate to ensure that the mechanism does not converge).

2. *For any stopping rule $\delta > 0$ there exists $\mu_0 \in \mathcal{A}$ such that the TFM mechanism terminates and $\mu_{(\mu_0,\delta)}$ is ϵ-stable, even when the TIM process does not converge.*

3. *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 for any stopping rule $\delta < \delta^*$ and any starting condition μ_0, $\eta(\Theta') < \epsilon$.*

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

III Empirical Application: The New South Wales Market

In this section we describe the details of the New South Wales college admissions system, use data from this market to show that students have "big-fish" peer preferences over a summary statistic of peer ability (**AA1**-**AA3**), and we justify our assumptions on program entry and exit (**AA4**-**AA6**).

III.A The New South Wales Tertiary Education Admissions System

Each state in Australia has a centralized body that processes college applications within its jurisdiction. Students in Australia 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." We study college admissions in New South Wales and the Australian Capital Territory[29] from 2003 to 2016.[30]

Students receive a score known as the *Australian Tertiary Admission Rank (ATAR)* which measures the student's academic percentile rank, over a re-normalized scale of 30-99.95. 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 (Manny, Yam, and Lipka, 2019). Therefore, we use the ATAR score as a proxy for student ability.

Each year, over 20,000 new high school graduates apply to programs where the ATAR score serves as a central admission criteria. To apply for admission, prospective students submit an ROL of up to nine programs to the United Admissions Centre (UAC), the centralized clearinghouse which processes applications to all major universities in NSW and the Capital Territory.[31] Students initially submit their ROLs before learning their own ATAR scores, but are able to costlessly change their ROLs after learning

[29] The Australian Capital Territory is a small, landlocked enclave surrounded by NSW and containing Australia's capital city Canberra.

[30] A number of changes to the matching process have occurred since 2016. Namely, students are now only able to list five programs on their ROL, and there is now a "guaranteed entry" option for students with ATAR above a particular threshold (Guillen et al., 2020).

[31] A minority of students, such as adult learners who do not have ATAR scores, apply directly to programs.

their ATAR score. Students are incentivized to submit initial ROLs early in the application process, as fees for stating initial ROLs increase over time.

Students and programs are matched using the student-proposing deferred acceptance mechanism which takes as inputs student ROLs, program rankings, and program capacities (Guillen et al., 2020).[32] Program rankings over students are determined by the sum of a student's ATAR score and program-student specific "bonus" points, which are awarded at the discretion of the program. Importantly, students can receive up to 10 bonus points at each program, and because bonus points are not observed by candidates before being matched, they serve as a significant source of admissions uncertainty.

The clearinghouse clearly informs students it is in their best interest to submit truthful ROLs:

> *"Your chance of being selected for a particular course is not decreased because you placed a course 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. Place the course you would like to do most at the top, your next most preferred second and so on down the list...If you're interested in several courses, enter the course codes in order of preference up to the maximum of nine course preferences."*[33]

The resulting matching mechanically creates a minimum ATAR score 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) at the program level every year. Going forward, we will refer to the clearly-in statistic for the cohort admitted in the previous year as the *Previous Year's Statistic (PYS)* for a particular program, and the clearly-in statistic for the current year as the *Current Year's Statistic (CYS)*.

When creating their ROLs, students do not know the CYS at any program. However, they can consult programs' PYSs as a guide–this information is made prominently available, by law, on the clearinghouse website (see Figure 1). Between 2003 and 2017–which contains our window of analysis–the only information about peer ability in the previous cohort revealed to applicants is the PYS.[34] For example, students applying for admission in 2016 are told the following:

> *"[PYS] for a course shows you the minimum selection rank needed by the majority of Year 12 applicants when offers were made in 2015. [CYS] for 2015–16 admissions won't be known until selection is actually made during the offer rounds. Use [PYS] as a guide when deciding on your preferences."*[35]

[32] Admissions take place in multiple rounds. We describe and analyze the process of the main round that takes place in early January, 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 that remain unmatched. As programs may elect not to enter subsequent rounds, there is a strong incentive for students to be matched to a desired program in the main round.

[33] See https://web.archive.org/web/20150918170643/http://www.uac.edu.au/undergraduate/apply/course-preferences.shtml , accessed 9/6/2021.

[34] After 2017, additional summary statistics about the previous year's ATAR distribution began to be disclosed.

[35] See https://web.archive.org/web/20150911225257/http://www.uac.edu.au/atar/cut-offs.shtml , accessed 9/6/2021.

As students do not know the number of bonus points they receive at each program, each student is uncertain ex ante about acceptance into a wide range of programs. Across programs, roughly half of all enrolling students have ATAR scores below the CYS of their program (Bagshaw and Ting, 2016), implying that the frequency of receiving bonus points is non-trivial.

III.B Data

We use data from the UAC clearinghouse for our analysis. Our data contain the universe of applications from graduating high schoolers processed by UAC for 2003-2016. We identify each student via a unique student id. Over this time period, there are on average 19 universities active per year, each offering numerous programs. We identify and track programs over time using a unique course code, and observe the program field of study.[36] For a subset of years (2010-2016) we observe students' ROLs at two points in time: immediately 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 our observation of these two ROLs. We observe the post-ROL for all years in our sample (2003-2016). In addition, we observe the students' 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 do not observe the final assigned program of each student. Unless otherwise specified, we use the sample of post-ROLs from 2003-2016 in our analysis.

III.C Applying Theory to Data: Assumptions and Identification Strategies

In this section, we discuss the assumptions specific to our empirical setting needed to identify peer preferences. A common and crucial step in identifying student preferences in market design research is assuming that submitted ROLs accurately reflect students' ordinal rankings over programs, and not strategic considerations to game the mechanism. **What we believe the submitted ROL reveals about students' true preferences is an essentially *untestable assumption*.** However, we can use strategic properties of the matching mechanism used, restrictions on our data, and our identification strategies, to support our assumptions. There are also stronger and weaker versions of this assumption, which lead, in our setting, to two entirely different identification strategies. We investigate both the strong and weak cases in the NSW data.

Both identification strategies rely on the assumption that students create or modify their ROL in part based on their ATAR scores relative to the PYSs of programs. Our setting does not allow us to examine whether students update their preferences in a "rational" manner when presented with this information,

[36] Prior to 2008, the same program could be listed twice according to its funding structure. The course code allows us to separately identify Commonwealth Supported Place (CSP) programs, which are subsidized, from Domestic Fee Paying (DFEE) programs. In 2008, all fee structures were standardized and all courses became CSP. In what follows, we treat DFEE courses as separate programs, but all of our results are robust to dropping DFEE programs.

or whether they over or under value aspects of it (as studied in Hastings, Van Weelden, and Weinstein, 2007). Nevertheless, our approaches address our central inquiry: whether student choice behavior depends on perceptions of relative peer abilities. Below we summarize how they impact our identification strategy, and in Section III.D we carry out each strategy. In Section III.D.4, we investigate and rule out alternative explanations of why student ROLs may depend on the relative values of ATAR scores and PYSs.

1. "Unconstrained" ROLs reveal students' rankings over programs given the PYS. In any strategy-proof mechanism, such as deferred acceptance, students have a weakly dominant strategy to report an ROL reflecting their true ordinal preferences over acceptable programs. In practice, there is often a cap placed by the market designer on the number of programs any student can list on their ROL, as is the case in our setting. However, students who have fewer acceptable programs than the cap retain a weakly dominant strategy to truthfully list their ROL, and those who have more acceptable programs than the cap will list the maximum number of allowable programs in any weakly undominated strategy (Haeringer and Klijn, 2009).[37] Therefore, a common approach is to view ROLs that list strictly fewer programs than the cap as accurately reflecting student preferences (Hastings, Kane, and Staiger, 2009; Abdulkadiroğlu, Agarwal, and Pathak, 2017; Luflade, 2019). In our data, 60% of students submit final ROLs that are "unconstrained"–i.e. list fewer than the maximum number of programs. Note that even if these students are not representative of those who list the maximum number of programs, the existence of peer preferences in a large subsample of the population can affect stability in the market.

2. Submitted ROLs reflect students' *relative* rankings over programs given the PYS. An emerging strand of the literature argues that the assumption that students play their weakly dominant strategy of truthfully submitting their ROL is too strong.[38] Fack, Grenet, and He (2019) argue students face a cost to reporting long ROLs, and therefore, if they have little uncertainty about their admission probability to any particular program (i.e. admission probabilities are sufficiently close to either zero or one), they may optimally omit "reach" and "safety" programs. Even in this case, an important result from Haeringer and Klijn (2009) still applies: the relative ranking of any two programs c and c' on a student's ROL will reflect her true relative ordinal preferences of c and c'. Under this weaker assumption, we look within person at how relative rankings over programs respond to new information about how a student's own ability measures up with peer quality. This identification strategy thus assumes that any changes between the pre-ROL and post-ROL reflect changes in a student's ordinal ranking of programs upon learning how her own ability matches up with peers at each program.

For a visual summary of our empirical approaches, see Figure A.1.

[37] This continues to hold in our setting with peer preferences under the ongoing assumption that students take the PYS as indicative of the CYS.

[38] See, for example, Chen and Sönmez (2006); Li (2017); Rees-Jones (2018); Sóvágó and Shorrer (2018); Chen and Pereyra (2019); Larroucau and Rios (2020); Artemov, Che, and He (2020); Hassidim, Romm, and Shorrer (2021).

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

Descriptive Evidence

Table 1 displays summary statistics of student ATAR scores, ROLs, program PYSs, and "score gaps," defined as the difference between the PYS of a ranked program and the student's ATAR, for the sample containing post-ROLs only (2003-2016), and for the sample containing pre- and post-ROLs (2010-2016). For both samples, we display program information both averaged across all programs on a student's ROL, and for the top-ranked program.

Table 1 indicates that the top-ranked program tends to have a higher PYS than programs ranked lower on student ROLs. This PYS is on average 7.6 points higher than the student's ATAR score. These statistics suggest that students have a general preference for higher quality programs, insomuch as the PYS is a signal of program quality. We provide more formal evidence to support this claim in Table A.3.

Figure 2 plots the proportion of top-ranked programs by score gap. Three clear patterns emerge. First, the upward sloping left-hand side of the graph suggests that students have a preference for "better" programs; the value of the horizontal axis is steeply increasing for programs with negative score gaps. If students' preferences were unrelated to program quality, we would not expect the proportion of top rankings to increase monotonically with the score gap.

Second, students do not want to be a "small fish" in their program of entry; the attenuated positive slope between a score gap of 1 and 6 suggests that students are trading off a preference for increased quality with a disutility for being overmatched by peers. After a score gap of 6, the dominant concern becomes relative peer ability.

Third, over 75% of students rank a "reach program" (defined as having a positive score gap) first on their post-ROL. Due to the support of possible bonus points, students have a non-zero admission probability to any program across the range of the x-axis of this figure. Recalling that no students in our sample have a binding constraint on the number of programs they are able to rank, the shape of this graph is not easily explained by considerations of admission probability on the part of students. In Section III.D.4 we discuss how the shape of this figure implies that students understand the strategic properties of the matching mechanism, nor is this shape consistent with other explanations presented in the literature.

We formally explore this pattern of "big-fish" preferences using two identification strategies. In all of our empirical analysis, we cluster standard errors at the program level. This is because all variation in the PYS occurs at the program level, and we study the effect of program PYS on student application behavior.

III.D.1 Across-Person Analysis

When creating their ROLs, students 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 student demand for a program this year? Under big-fish preferences, students will be less likely to rank

programs with PYSs that are further above their own ATAR scores.

We test for this using changes in programs' PYSs across time. 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 student score, the number of students who apply, the 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 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 students?

The results are presented in Table 2 and support our theory of "big-fish" preferences. When a program's PYS increases by one point, fewer (column 2) students rank the program on their ROLs, and these students tend to be higher scoring (column 1). Columns 4 and 5 test these effects discontinuously – the dependent variable splits the sample into students with ATAR scores either above or below the PYS of program c, and quantifies the percentage who have ranked that program. These two columns show that the effects in Columns 1 and 2 are driven primarily by those with scores *below* the PYS, who become less likely to rank the program on their final ROL.

This test assumes that students react to the PYS due to peers, and not to changes in program quality. To test this assumption, we re-run the specification while including lagged values of a programs' PYS in Table A.4. Lagged values of the PYS (from two and three years before the current year) have little predictive power, indicating that responses to the PYS are not based on a trend of changes in the PYS over time.

III.D.2 Within-Person Analysis

We next measure how students respond to new information about their own relative ability. We observe students' ROLs at two points in time: both before and after they learn their final ATAR score. Students are incentivized to submit preferences early through lower application fees, before learning their ATAR score, and the overwhelming majority (99.5%) of students do so. However, they can update their ROL after learning their score. If students have big-fish peer preferences, then learning their own relative score will impact their preferences. Students will deprioritize programs with PYSs that far exceed their own ATAR score.

Our strategy for identifying peer preferences leverages not only *that* students adjust their ROLs after learning their ATAR scores, but also *how* they adjust their ROLs. 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*.

Switches are particularly difficult to explain unless students' preferred programs change after observing their ATAR scores. While additions and removals can be justified by low assessed probability of enrollment in a program, switching the relative ranking of two programs is not easily explained by probability of admissions calculations, and implies that either the pre- or post-ROL is weakly dominated.

Appendix Table A.5 shows 46% of students do not submit the same pre- and post-ROL. On average, each student makes 1.93 adjustments, which corresponds to 29% of the pre-ROL. Of the adjustments made, switches are the most common.

Adjustments that students make result in a smaller PYS/ATAR score gap, that is, the difference between the program PYS and the student's ATAR score. We graphically present evidence of how switches affect the PYS/ATAR score gap.Figure 3 considers students who switch the ranking of the program initially listed first on their pre-ROLs. It plots the students' ATAR scores against the PYS/ATAR score gap for top-ranked programs, and finds that the PYS/ATAR score gap 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, top-ranking programs with PYS<ATAR on their pre-ROLs.

We quantitatively analyze changes to the ROLs through linear regressions. Specifically, we regress indicators $y_{c,t,\theta}$ for whether a program c was removed, added, or promoted by student θ in year t on the PYS/ATAR score gap between c and θ.[39] We run the following regressions

$$y_{c,t,\theta} = \beta(PYS_{c,t} - ATAR_\theta) + \alpha_c + X_\theta + \epsilon_{c,t,\theta} \tag{2}$$

$PYS_{c,t} - ATAR_\theta$ is student θ's score gap at program c in year t, α_c represents a program fixed effect, and X_θ represents a vector of pre-ROL characteristics for student θ (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 3 displays the results from this regression for the "promote" variable, and we show the results for "add" and "remove" in the appendix.[40] All three support the presence of "big-fish" preferences. Pro-

[39]The variable "remove" (and "add") indicate that a program is removed from (or added to) a student's pre-ROL after learning her 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-ROL than on the pre-ROL, ignoring all other adds and drops. To define promotion, we use the following inversion algorithm: First, keep only programs that are on both the pre- and post-ROL (i.e. remove all adds and removals from both lists), and call these the redacted pre- and post-ROLs, respectively. Second, define a program as promoted if it is ranked in a higher spot on the redacted post-ROL than on the redacted pre-ROL. Because any switch results in one program being promoted and one program being demoted, we do not include "demote" in our regression analysis.

[40]The sample for these tables is all students who rank at most eight programs in both their pre- and post-ROLs. Following our identifying assumptions, we do not need this restriction for the "promote" regressions, and results are similar without it.

grams that are added or promoted within the ROL have PYSs that are systematically lower than those that are removed, and are closer to the student's ATAR score. Programs that are removed have a larger score gap, that is, they are more of a "reach" program. Specifically, over two-thirds of programs dropped have positive score gaps, and there are more programs dropped with a positive score gap of strictly less than 10 (and for which admissions probabilities are strictly positive) than those dropped with a negative score gap. This asymmetry matches the switching pattern we observe; students are more likely to reprioritize programs with PYSs below their own ATAR score than those with PYSs above their ATAR score.

These effects persist with an array of fixed effects. In Columns 3-6 we attempt to compare the behavior of students who construct very similar pre-ROLs by including fixed effects for ROLs of the same length or ranking the same programs at the top of the ROL. For example Column 6 provides evidence of how the adjustments made by students whose pre-ROLs have the same three programs ranked in the top three spots (in the same order) 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."

An important assumption is that pre-ROLs reflect student preferences (given initial beliefs), instead of mere placeholders. We believe this assumption is justified. First, while there is no monetary cost to changing ROLs, there is an administrative cost to doing so, and therefore, students minimize costs by not arbitrarily constructing the pre-ROL. Second, there is a high correlation between students' pre- and post- ROLs, as seen in Appendix Table A.5, which also suggests that the pre-ROL is not constructed arbitrarily. Finally, changes (especially switches) to a student's ROL are predicted by the difference between the realization of their PYS of a program and their own ATAR score, which is not known at the time the pre-ROL is created. If the pre-ROL were arbitrary, we would expect little correlation between the score gap and ROL changes.

III.D.3 How Important Are Peer Preferences?

How much do changes in students' ROLs affect the final matching? This question is important for at least two reasons. First, a large effect lends credence to our identification strategies. Recent work by Artemov, Che, and He (2020) suggests that students may include arbitrary information on their ROLs that do not (with high probability) affect the final matching; for example, a student may rank "reach" programs she is unlikely to be admitted to in arbitrary order.[41] As argued by Artemov, Che, and He (2020), we should not view these changes to the ROL as arbitrary if they affect the final matching.

[41] This is motivated by Sóvágó and Shorrer (2018), Artemov, Che, and He (2020), and Hassidim, Romm, and Shorrer (2021) who study "obviously dominated" choices: in some markets in which students can apply to each program with or without funding, some students rank the unfunded version of the program above the funded version of the program, or fail to rank the funded version at all. Artemov, Che, and He (2020) find that the vast majority of these such oddities do not affect the final matching.

Second, a large effect of changes to the pre-ROL on the final matching suggests that peer preferences have a large overall impact, and are an especially important consideration for effective market design.

Changes to the pre-ROL have a large effect on the final matching. Figure 4 plots students' 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).[42] The approximate share of students who would have received a different final matching under their pre-ROL than under their post-ROL is the sum of the absolute value of the difference between the red and blue dots for each preference number.[43] The post-ROL increases the probability of getting one's first-choice program by 22%, and of receiving a different final matching by 25%. Recalling that 46% of students have different pre- and post-ROLs, the percent of students who change their ROL who then receive a different matching than they would have under their pre-ROL is approximately 54%. We find evidence that switches, the most difficult adjustment to account for via traditional models, are even more payoff relevant than other adjustments; when we restrict our sample to students who only make switches to their pre-ROL (i.e. do not add or remove and programs), we estimate an even higher fraction of students would have been matched to a different program under their pre-ROL.

III.D.4 Alternative explanations

Our empirical findings are consistent with the existence of peer preferences. Could other reasonable models generate similar effects in our data? In Appendix F we evaluate the following models:

- Students cannot reason through optimal behavior in the matching mechanism (Li, 2017),
- Student have other non-classical preferences, including loss aversion, which makes them disprefer rejection (Dreyfuss, Heffetz, and Rabin, 2021; Meisner and von Wangenheim, 2019; Meisner, 2021),
- Students are uncertain about their preferences over programs, and will optimally acquire more (costly) information over programs to which they have a higher chance of admission (Grenet, He, and Kübler, 2022; Immorlica et al., 2020; Hakimov, Kübler, and Pan, 2021). Risk averse students are more likely to highly rank programs for which they have gathered information,
- Students are wary of their "fit" or "mismatch" with a program (Rothstein and Yoon, 2008; Conger, Long, and McGhee Jr, 2020), and use a program's PYS as a signal of fit.

In a matching market the size of that in NSW, it is likely that there are some students whose preferences reflect each of these models; however, we do not find strong evidence to support these

[42] For robustness, we repeat this exercise but vary how we calculate admissions probabilities. As admission is determined by 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. Figure 4 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.

[43] This share is approximate because the length of each student's pre- and post-ROLs are not necessarily equal. However, as seen in Table 1, the length of the pre- and post-ROLs are similar for most students.

alternatives as main drivers for the effects we observe. As we derive in the appendix, most of these alternative models imply that lower admissions chances make ranking a program less desirable. Because students have a lower probability of being admitted to a program with a PYS that just exceeds their ATAR score (i.e. there is strictly positive probability of receiving zero bonus points), these models therefore predict a "missing mass" of students who rank programs with a PYS just exceeding their ATAR scores first on their post-ROL. Empirically, we observe no such missing mass, indeed students are more likely to top-rank programs whose PYSs slightly exceed their own ATAR scores (see Figure 2). Moreover, these alternatives similarly imply a discontinuously larger fraction of students whose top-ranked program on their pre-ROL has a PYS just exceeding their ATAR score to demote this program compared to students whose top-ranked program on their pre-ROL has a PYS that is just exceeded by their ATAR score. We see no such difference in Figure 3; students with ATAR scores of 86 have an average score gap of zero, and we see similar adjustments to those with ATAR scores just above and below this level. Therefore, these alternatives do not appear to explain our documented results in aggregate.

III.E Empirical Evidence of Changes in the Market

In Section II.E.1 we assume in our analysis that the set of programs changes from year to year, but that other factors (for example, the popularity of certain fields of study or the popularity of certain universities) do not. We test this by looking at the distribution of additional variables (other than the PYS) over time. Figure A.2 shows that aggregate student preferences over fields of study and university campuses 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.E.1 assumes that entry and exit of programs occurs for programs that have a low PYS. Figure 5 shows the somewhat bimodal distribution of program "age"; while some programs exist for 14 or more years, the majority exist for fewer than four years. In addition, we plot the difference in PYS for each age cohort relative to the youngest programs. Programs that enter and exit frequently have lower PYSs than incumbents. In the following section, we investigate the impact of entry and exit of less popular programs on stability.

IV Empirical Evidence of Instability in NSW

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 6 provides evidence of convergence. Panel 1 plots the interquartile range of the PYS by year, for programs with PYSs between 65 and 75 in 2012, while Panel 2 plots the interquartile range of the PYS by year, for 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 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 less variable than in years immediately preceding 2012. If these plots were driven largely by mean reversion of the PYSs in the base year, we would not expect the PYSs in years 2013-2016 to remain nearly constant.

Figure 7 displays the intensive and extensive margins of change in PYS from year to year by program age. Our data allow us to track programs for up to 14 years. In Panel 1, we observe that the absolute value of the change in PYS is decreasing over time. Programs initially have an average year-to-year change of 2.3 points, which is an economically meaningful magnitude; this equals 26% of a standard deviation of the distribution of program PYSs ranked across students (see Table 1). Programs in years 12-14 have an average year-to-year change in their PYS of half a point. Figure A.3 recreates this visualization and controls for entry and exit of programs into our data by grouping programs by the number of years each program is observed in our data. Across all groups, we observe a similar decreasing trend in the absolute change in PYS over time, which falls to under 1 point as programs age beyond 10 years (four of five of the groups in our data for at least 10 years have all point estimates beyond year 10 less than 1 point.) Panel 2 shows that older programs are also more likely to have no change in PYS from the year before, i.e. the CYS is equal to the PYS. Initially, 27% of programs experience no change in PYS year-over-year, but this number climbs to over 60% for programs in years 12-14.

While the PYSs of individual programs converge over time, entry and exit prevents the PYSs of short-lived programs from reaching (near) steady state.

To discuss the impact of this instability, we focus on a related outcome: attrition. We define attrition as occurring when a student neither graduates from their matched program, nor returns in the following year. Whenever a blocking pair is consummated (either with a different program or the student's outside option), attrition occurs, and therefore, we expect attrition to be higher at programs with students who have more blocking pairs. For privacy reasons, we do not observe attrition at the individual level, but instead merge in the attrition rate at the university-year level. Remark 7 in the appendix (an extension of Theorem 4) predicts that students at programs with larger, positive changes in the PYS are more likely to be in blocking pairs. Intuitively, these students are those who are unexpectedly surprised to find that they are overmatched by their peers, and suffer big-fish losses to utility; our model predicts that only students at these programs form "negative utility" blocking pairs and prefer being unmatched to remaining at their current program. Table 4 shows that, at the program level, higher yearly changes in the PYS are correlated with higher attrition rates. This relationship is positive and significant; a 1 point increase in CYS-PYS is associated with a 2.8% increase in the attrition rate of a program. This pattern is robust to controls for year, field of study, and program age.

We discuss in the appendix how students from low-socioeconomic backgrounds, aboriginal students,

and students with disabilities are more likely to attend programs with large changes in PYS where attrition is high.

V Conclusion

How important is it that a matching market allows agents to fully express their preferences? We study this question in markets in which students have preferences over their peers which cannot be directly expressed to the matching mechanism. We show that a dynamic process which reveals the composition of previous cohorts can lead to a stable matching in the long run, forming a tâtonnement process.

Using data from the New South Wales college admissions market, we provide evidence for the existence of "big-fish" peer preferences; students prefer not to attend programs where they are overmatched by peers. The functional form of peer preferences can vary across education markets, and we provide a simple empirical test for stability regardless of the form peer preferences take. This test can be applied without detailed information of peer preferences, or detailed application data.

We use our test for stability to show that long-lived programs in the NSW market converge to stability. This matches our theoretical analysis which finds that key features of the NSW market guarantee this convergence for long-lived programs, but that entry and exit of programs causes instability for students matched to short-lived programs. This instability is correlated with higher attrition among affected students. Moreover, the failure to explicitly design the market to account for peer preferences bears an unequal cost on students of different demographic groups: we discuss in Appendix G that low-socioeconomic status students, aboriginal students, and students with disabilities are particularly likely to be affected by this instability.

A static mechanism that delivers a stable matching is likely infeasible as it requires soliciting functional preferences from students. We propose a new mechanism that induces an iterative process *within each cohort* and is a relatively small modification to iterative mechanisms already in use in higher education markets in China, Brazil, Germany and Tunisia (see Bo and Hakimov (2019); Luflade (2019)). This mechanism removes the sources of instability in the NSW market.

Disclosing more detailed information of the composition of previous cohorts may be an additional design decision that is important in fully ensuring stability. In Appendix E, we show that student preferences may not be accurately captured via a summary statistic of the distribution of student types. That the NSW market reveals only a summary statistic of the abilities of students (the PYS) potentially limits our ability to observe nuances of the functional form of peer preferences, and potentially limits the ability of students to accurately select their most desired programs, and may lead to instability. We support recent proposals aimed at reporting additional details of the ATAR distribution within program to applicants.[44]

[44]For more information, see https://www.teqsa.gov.au/latest-news/publications/improving-transparency-higher-education-admissions.

A question remains. What causes peer preferences? Peer preferences could be caused by a direct aversion to being a "small fish in a big pond," or they could be signs of market failures that can be addressed by market design solutions. For example, if enrollment in individual classes is determined by class rank, a student may avoid a preferred program in order to ensure herself a desirable course schedule. A market that resolves course allocation ex ante may reduce the magnitude of "peer preferences" in the match. Studying these microfoundations is left for future research.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A. Pathak. 2017. "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match." *American Economic Review* 107 (12):3635–3689.

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag A. 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.

Ainsworth, Robert, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola. 2020. "Information, Preferences, and Household Demand for School Value Added." NBER WP 28267.

Allende, Claudia. 2020. "Competition Under Social Interactions and the Design of Education Policies." Mimeo.

Artemov, Georgy, Yeon-Koo Che, and YingHua He. 2020. "Strategic 'Mistakes': Implications for Market Design Research." Mimeo.

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

Avery, Christopher and Parag A. Pathak. 2021. "The Distributional Consequences of Public School Choice." *American Economic Review* 111 (1):129–152.

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.

Azmat, Ghazala and Nagore Iriberri. 2010. "The importance of relative performance feedback information: Evidence from a natural experiment using high school students." *Journal of Public Economics* 94 (7):435–452.

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

Balinski, Michel and Tayfun Sönmez. 1999. "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory* 84:73–94.

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

Beuermann, Diether W. and C. Kirabo Jackson. 2019. "The Short and Long-Run Effects of Attending The Schools that Parents Prefer." NBER WP 24920.

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." Mimeo.

Bo, Inacio and Rustamdjan Hakimov. 2019. "The iterative deferred acceptance mechanism." Mimeo.

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. 2021. "Can Market Participants Report their Preferences Accurately (Enough)?" *Management Science, Forthcoming* .

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." Mimeo.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3):855–882.

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. 2006. "School Choice: An Experimental Study." *Journal of Economic Theory* 127 (1):202–231.

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

Conger, Dylan, Mark C Long, and Raymond McGhee Jr. 2020. "Advanced Placement and Initial College Enrollment: Evidence from an Experiment. EdWorkingPaper No. 20-340." *Annenberg Institute at Brown University* .

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* .

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2019. "School Starting Age and Cognitive Development." *Journal of Policy Analysis and Management* 38 (3):538–578.

Dillon, Eleanor Wiske and Jeffrey Andrew Smith. 2017. "Determinants of the Match between Student Ability and College Quality." *Journal of Labor Economics* 35 (1):45–66.

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. 2021. "Expectations-Based Loss Aversion May Help Explain Seemingly Dominated Choices in Strategy-Proof Mechanisms." *American Economic Journal: Microeconomics, Forthcoming* .

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." Mimeo.

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

Esponda, Ignacio and Demian Pouzo. 2016. "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models." *Econometrica* 84 (3):1093–1130.

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.

Feather, Norman T. 1989. "Attitudes towards the high achiever: The fall of the Tall Poppy." *Australian Journal of Psychology* 41 (3):239–267.

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.

Greinecker, Michael and Christopher Kah. 2021. "Pairwise stable matching in large economies." *Econometrica* 89 (6):2929–2974.

Grenet, Julien, YingHua He, and Dorothea Kübler. 2022. "Preference Discovery in University Admissions: The Case for Dynamic Multi-offer Mechanisms." *Journal of Political Economy, Forthcoming* .

Grigoryan, Aram. 2021. "School Choice and the Housing Market." Mimeo.

Guillen, Pablo, Onur Kesten, Alexander Kiefer, and Mark Melatos. 2020. "A Field Evaluation of a Matching Mechanism: University Applicant Behaviour in Australia." *The University of Sidney Economics Working paper Series* .

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

Hakimov, Rustamdjan, Dorothea Kübler, and Siqi Pan. 2021. "Costly information acquisition in centralized matching markets." Mimeo.

Hassidim, Avinatan, Assaf Romm, and Ran I. Shorrer. 2021. "The Limits of Incentives in Economic Matching Procedures." *Management Science* 67 (2):951–963.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2009. "Heterogeneous Preferences and the Efficacy of Public School Choice." Mimeo.

Hastings, Justine S., Richard Van Weelden, and Jeffrey Weinstein. 2007. "Preferences, Information, and Parental Choice Behavior in Public School Choice." NBER WP 12995.

Immorlica, Nicole S., Jacob D. Leshno, Irene Y. Lo, and Brendan J. Lucier. 2020. "Information Acquisition in Matching Markets: The Role of Price Discovery." Mimeo.

Klaus, Bettina and Flip Klijn. 2005. "Stable matchings and preferences of couples." *Journal of Economic Theory* 121 (1):75–106.

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.

Larroucau, Tomás and Ignacio Rios. 2020. "Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem." Mimeo.

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

Li, Shengwu. 2017. "Obviously strategy-proof mechanisms." *American Economic Review* 107 (11):3257–87.

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

Manny, Anthony, Helen Yam, and Robert Lipka. 2019. "The Usefulness of the ATAR as a Measure of Academic Achievement and Potential." *https://www.uac.edu.au/assets/documents/submissions/usefulness-of-the-atar-report.pdf* .

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.

Meisner, Vincent. 2021. "Report-dependent utility and strategy-proofness." Mimeo.

Meisner, Vincent and Jonas von Wangenheim. 2019. "School choice and loss aversion." Mimeo.

Moschovakis, Yiannis. 2006. *Notes on Set Theory, Second Edition*. Springer.

Murphy, Richard and Felix Weinhardt. 2020. "Top of the Class: The Importance of Ordinal Rank." *Review of Economic Studies* 87 (6):2777–2826.

Nei, Stephen and Bobak Pakzad-Hurson. 2021. "Strategic Disaggregation in Matching Markets." *Journal of Economic Theory* 197.

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

Nguyen, Thanh and Rakesh Vohra. 2018. "Near-feasible stable matchings with couples." *American Economic Review* 108 (11):3154–69.

Pathak, Parag A. and Tayfun Sönmez. 2008. "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review* 98 (4):1636–1652.

Pop-Eleches, Cristian and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4):1289–1324.

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." Mimeo.

Qiu, Junping and Rongying Zhao. 2007. *College admissions cutoffs and application guide: 2007-2008, Second Edition*. Science Press: Beijing.

Rees-Jones, Alex. 2018. "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match." *Games and Economic Behavior* 108:317–330.

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 and Albert Yoon. 2008. "Mismatch in law school." NBER WP 14275.

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. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly Journal of Economics* 116 (2):681–704.

———. 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.

———. 2014. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics* 6:253–272.

Sasaki, Hiroo and Manabu Toda. 1996. "Two-Sided Matching Problems with Externalities." *Journal of Economic Theory* 70 (1):93–108.

Scarf, Herbert. 1960. "Some examples of global instability of the competitive equilibrium." *International Economic Review* 1 (3):157–172.

Scotchmer, Suzanne and Chris Shannon. 2015. "Verifiability and group formation in markets." Mimeo.

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.

Song, Yan, Kentaro Tomoeda, and Xiaoyu Xia. 2020. "Sophistication and Cautiousness in College Applications." Mimeo.

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.

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

Teske, Paul, Jody Fitzpatrick, and Gabriel Kaplan. 2007. "Opening Doors: How Low-Income Parents Search for the Right School." Tech. rep., University of Washington, Daniel J. Evans School of Public Affairs.

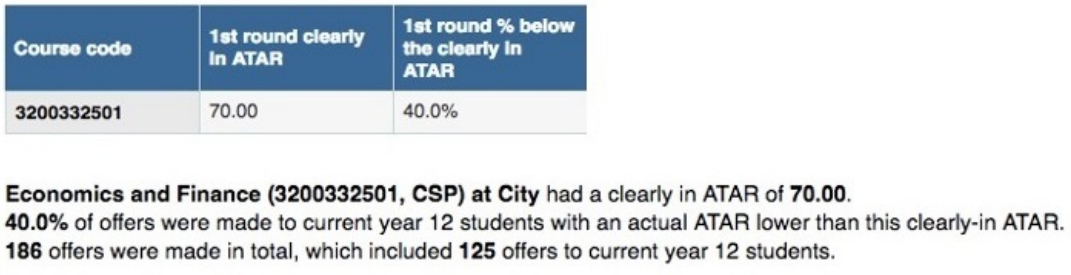
Tincani, Michela M. 2018. "Heterogeneous Peer Effects in the Classroom." Mimeo.

Tran, Anh and Richard Zeckhauser. 2012. "Rank as an inherent incentive: Evidence from a field experiment." *Journal of Public Economics* 96 (9):645–650.

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." Mimeo.

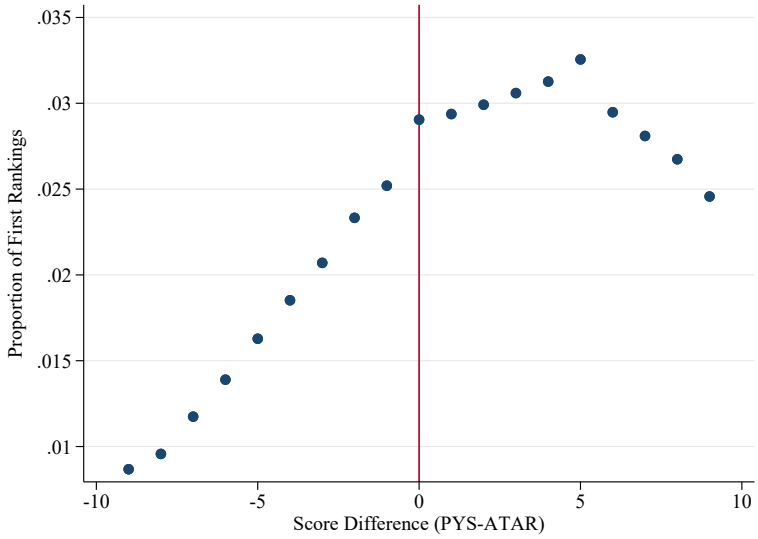
Figure 1: Example of Information Provided to Students about a Program's PYS



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

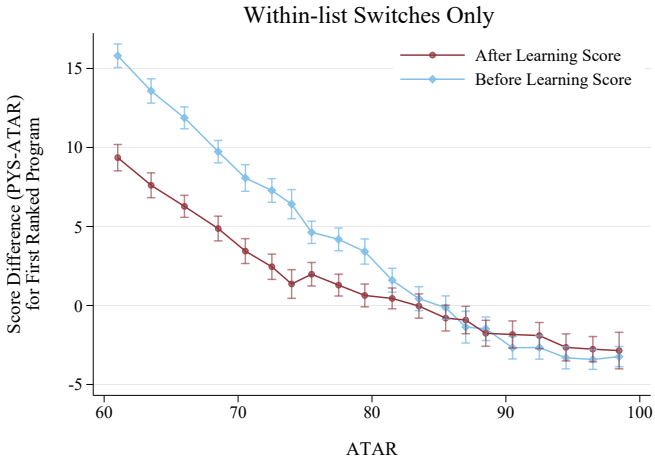
Economics and Finance (3200332501, CSP) at City had a clearly in ATAR of **70.00**.
40.0% of offers were made to current year 12 students with an actual ATAR lower than this clearly-in ATAR.
186 offers were made in total, which included **125** offers to current year 12 students.

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



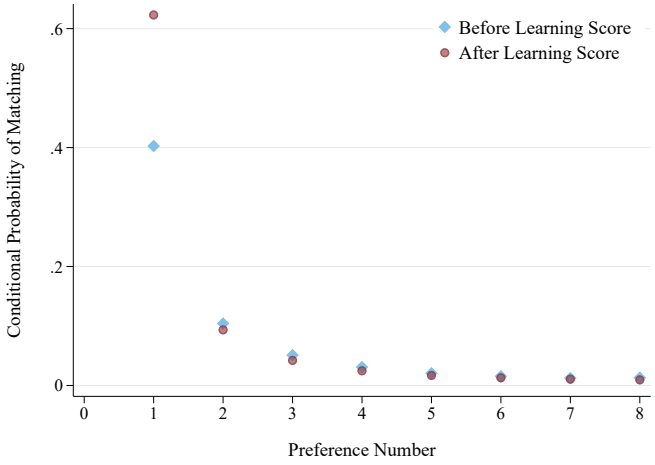
This figure studies the top-ranked programs listed by students after they learn their ATAR score. On the x-axis is the gap between the top-ranked program's PYS and the student's ATAR score. Because students can receive up to 10 bonus points, there is a positive probability of admission to all programs with score gaps in the presented range of the x-axis. On the y-axis is the proportion of all top-ranked programs that have that score gap The off-center, single-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. That the graph is increasing until a positive score difference of 6 point, suggests that students are more likely to rank "better," high-PYS programs, even if they are slightly overmatched by peers. There is no discontinuity in the figure to the right of 0, which we would expect to occur if students misunderstood the mechanism, as we discuss in Section III.D.4. The downward slope for score differences greater than 6 suggests that while students are not afraid to rank "reach" programs, they become gradually less attractive as the score gap increases.

Figure 3: Average Listed Program PYS before and after Score Revelation, Restricting to Switched Programs (first ranking only)



This figure plots the average score gap between the PYS for the top-ranked program listed by student ROLs and that student's ATAR score. It displays the gap for top-ranked programs on the pre- (blue) and post-ROLs (red), restricted to the set of students whose top choice on the pre-ROL also appears on the post-ROL. Lower-scoring students rearrange their lists to prioritize lower PYS programs, which higher-scoring students rearrange their lists to prioritize higher PYS programs. We use the pre- and post-ROL sample from 2010-2016. The two plots appear to cross near an ATAR of 85, which corresponds to a zero score gap on average on the pre-ROL. 95% confidence intervals using standard errors clustered at the program level are indicated.

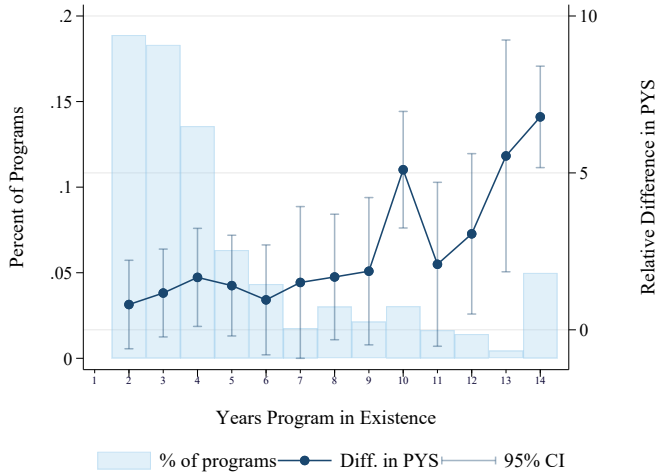
Figure 4: Conditional Probability of Matching before and after learning Score, by Rank Order



We calculate a student's probability of matching to each program on her post-ROL. We do this calculation for both pre- and post-ROLs. 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 ranked 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. The approximate 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 post-ROL is

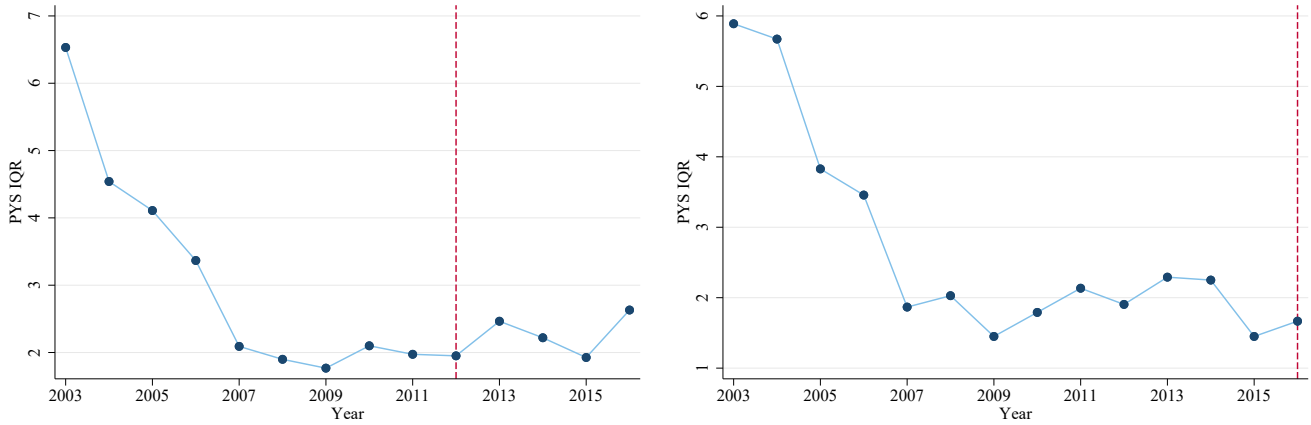
$$\sum_{j=1}^{8} \Pr(\text{Matched to preference number } j \text{ in post-ROL}) - \Pr(\text{Matched to preference number } j \text{ in post- ROL } \textit{if instead submitted pre-ROL}).$$

Figure 5: Difference in PYS by length of program existence



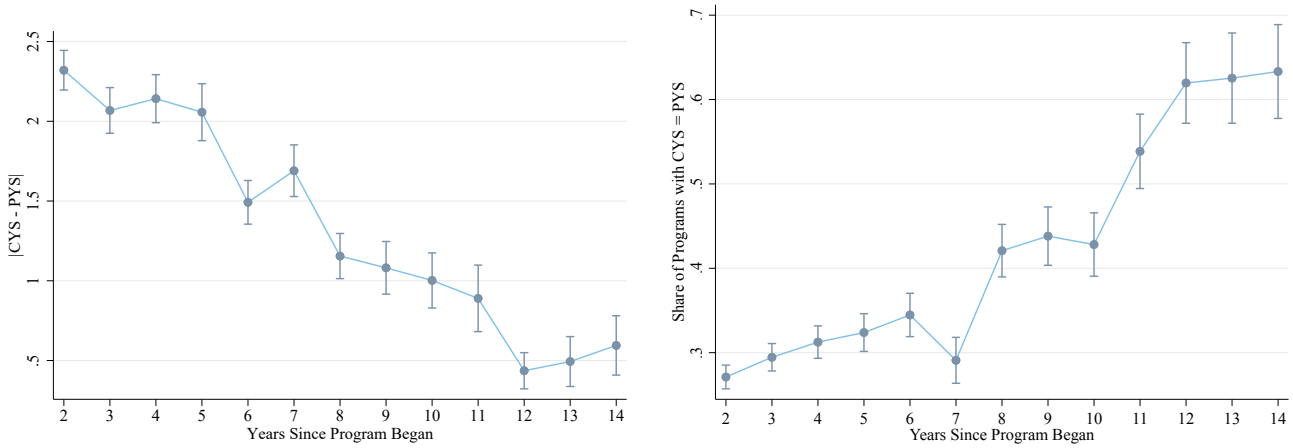
This figure plots the estimated difference in PYSs for programs based on how long they exist in our data, relative to programs that exist for only one year. These estimates control for the initial calendar year in which the program enters, and the field of study. Programs that exist for 14 years, for 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. 95% confidence intervals using standard errors clustered at the program level are indicated.

Figure 6: Convergence test for programs with ultimately similar PYSs



These figures provide evidence for convergence of PYS within program over time. The left 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 backward 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 right 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.4) 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.

Figure 7: Year-to-Year Variation in PYS Within Program, Over Time



These figures provide evidence for convergence of PYS within program over time. The left figure plots $\Delta_{c,t} := |CYS_{c,t} - PYS_{c,t}|$, a measure of PYS "instability," averaged over programs against the number of years the program has existed. This average appears to converge to zero over time, which is consistent with the PYS of each program reaching steady state. On the right, we plot the share of programs for which the CYS is equal to the PYS – i.e. the share of programs with $\Delta_{c,t} = 0$. This share increases with the age of the programs. Note that any variation in the market or applicant pool from year to year may lead to small deviations in the CYS, preventing this share from reaching 1. 95% confidence intervals using standard errors clustered at the program level are indicated.

Table 1: Student and ROL Summary Statistics

Variable	Obs	Mean	Std. Dev.	P25	P50	P75
Pre- and Post-ROL Sample (2010-2016)						
Student ATAR Score	104,519	72.1	18.6	59	75	88
# of Programs Ranked	104,519	5.7	1.9	4	6	7
All Programs in a Student's ROL						
Avg. PYS	104,519	78.8	9.2	72	78.5	85.8
Avg. Pre-ATAR PYS	103,375	79.3	9.1	72.5	79	86.2
Avg. Score Gap	104,519	6.7	14.4	-3.2	2.9	14.4
Avg. Pre-ATAR Score Gap	103,375	7.1	14.7	-3.4	3.7	15.5
Only Top-Ranked Program in a Student's ROL						
PYS	88,769	80.4	11.6	71.8	80.1	90.3
Pre-ATAR PYS	87,692	81.1	11.3	72.6	81	91
Score Gap	88,769	7.8	14.1	-1	4.7	14.7
Pre-ATAR Score Gap	87,692	8.9	14.9	-1	6	17
Post-ROL Sample (2003-2016)						
Student ATAR Score	289,500	72.9	18.3	60	76	88
# of Programs Ranked	289,500	5.6	1.9	4	6	7
All Programs in a Student's ROL						
Avg. PYS	289,500	78.6	9	72	78.2	85.3
Avg. Score Gap	289,500	5.8	14.2	-3.8	1.9	13.1
Only Top-Ranked Program in a Student's ROL						
PYS	243,195	80.4	11.3	71.3	80.1	90
Score Gap	243,195	7.6	14.2	-1	4.1	14.1

This table displays summary statistics on student ATAR scores, Score Gaps (program PYS minus student ATAR score), ROLs, and associated program PYSs for students who rank at most 8 programs on any observed ROL within the respective sample. Rows 3-6 and 13-14 examine the average PYS for *all* programs listed by a student, whereas rows 7-10 and 15-16 focus only on the top-ranked program. The primary reason for variation in observation counts within sample is that the PYS is not defined for a program in its first year of existence.

Table 2: Across Time Student Response to Program PYS

	(1) Avg. Student Score	(2) # of Students	(3) % of Students	(4) % of Students Higher Score	(5) % of Students Lower Score
Previous Year Statistic	0.344***	-2.650***	-0.008***	-0.003	-0.015***
	(0.015)	(0.262)	(0.001)	(0.001)	(0.001)
Observations	14,853	14,853	14,853	14,853	14,853

This table shows the estimated β coefficients of equation (1) were $y_{c,t}$ is the average student score, the number of students who apply, the 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 the program level. $^{*}p<0.05,^{**}p<0.01,^{***}p<0.001$

Table 3: Impact of Score Gap on Promote

	(1) Promote	(2) Promote	(3) Promote	(4) Promote	(5) Promote	(6) Promote
PYS - ATAR	-0.0017***	-0.0015***	-0.0017***	-0.0016***	-0.0014***	-0.0012***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Program FE		✓				
ROL length FE			✓			
Top Progam FE				✓		
Top 2 Programs FE					✓	
Top 3 Programs FE						✓
Observations	537,442	537,406	537,442	537,442	537,442	537,442

The dependent variable is an indicator for whether a program was promoted from a student's pre-ROL to the post-ROL. Column (2) includes program fixed effects, column (3) includes a fixed effect for the number of programs listed on a student's pre-ROL, column (4) includes a fixed effect for the top-ranked program in the pre-list, column (5) includes a fixed effect for the top two ranked programs in the pre-list, column (6) includes a fixed effect for the top three ranked programs in the pre-list. We use the pre- and post-ROL sample from 2010-2016. Standard errors in parentheses, clustered at the program level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 4: Relationship Between Attrition Rate and CYS-PYS

Program-years with CYS-PYS > 0 (*N = 4,094*)			
CYS-PYS	0.327***	0.314***	0.261***
	(0.023)	(0.024)	(0.024)
Means: Attrition Rate = 11.6; |CYS-PYS| = 2.5			
Year FE	✓	✓	✓
Field FE		✓	✓
Course Age FE			✓

This table tests for the relationship between the year-to-year change in PYS of a given program and its attrition rate (measured in percentage points). Following Remark 7, we focus on program-years with CYS>PYS. We find that, across a host models with various fixed effects, programs with more volatility in their PYS also have higher student attrition rates. Standard errors in parentheses, clustered at the program level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

APPENDIX: FOR ONLINE PUBLICATION

Natalie Cox Ricardo Fonseca Bobak Pakzad-Hurson

This document presents examples and proofs omitted in the main text, additional theoretical results, and additional empirical evidence.

A Examples discussed in main text

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

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

Each θ receives utility $u^\theta(c|\alpha) = v^\theta - f(s(\lambda(\alpha)), r^\theta)$ from attending program c given α, where

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

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 below average at the program.[1] Therefore, each θ is better off enrolling at c if and only if $v^\theta - f(s(\lambda(\alpha)), r^\theta) \geq 0$, where we break ties in favor of the student attending the program. Let each v^θ be distributed independently and uniformly over $(0,1)$.

Initialize the TIM process with μ_0 such that $s(\lambda(\mu_0)) \leq 1-q$ and let $s_t = s(\lambda(\mu_t))$ for $t \geq 0$. 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 scoring measure of students enrolls at c at $t=1$ because they expect (mistakenly for some) to face no peer loss from doing so.

What about (p_2, s_2)? Given s_1, only the $1-k$ fraction of students with $r^\theta < s_1$ for whom $v^\theta \geq k$ prefer c to remaining unmatched, and all students with $r^\theta \geq s_1$ prefer c to being unmatched.

We consider the case in which the program fills all of its seats in μ_2, i.e. $p_2 = 1 - \frac{q}{2} - \frac{q}{2(1-k)}$. This 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 $k \geq \frac{4}{5}$, $s_2 \leq 1-q$.[2] But note then that $(p_3, \lambda_3) = (p_1, \lambda_1)$, as now all students with scores $r^\theta > 1-q$

[1] Note that $f(\cdot, r^\theta)$ is not continuous in its first argument. We make two observations. First, this market still satisfies condition **A4**, implying by Theorem 1 that a stable matching exists. Second, the qualitative results of this example would be unchanged if we replaced $f(\cdot, r^\theta)$ with a function that is continuous, but with a steep slope around the point where $s(\lambda(\alpha)) = r^\theta$, but the current example leads to cleaner calculations.

[2] 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 $q \leq \frac{1}{3}$.

wish to attend c given s_2. 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 market 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 $(s_* - p_*)(1-k) = q - (1 - s_*)$. 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 these equations yields:

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

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 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.

We now consider an example that is nearly identical to Example 1, 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 2. *Consider Example 1 but where $s(\lambda(\alpha))$ represents the median of scores r^θ of enrolled students at the program given α, that is, $s^c(\lambda(\alpha)) = \sup\{r | \frac{\lambda^{c,(r,r)}(\alpha)}{\lambda^{c,(1,1)}(\alpha)} \leq \frac{1}{2}\}$.*

*The cutoff and median score at $t = 1$ remains the same as in Example 1, 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. Therefore, $\succeq^{\theta|\mu_1} = \succeq^{\theta|\mu_2}$ for all $\theta \in \Theta$. By Assumption **A1** it must be that $\mu_2(\theta) = \mu_3(\theta)$ for almost all $\theta \in \Theta$. Therefore, $\lambda(\mu_2) = \lambda(\mu_3)$ and by Theorem 2, the TIM process produces a stable matching for all $t \geq 2$.*

B Proofs

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:

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

and the second given by:

$$Z^\lambda(p,\lambda) = \lambda(A(p,\lambda)). \tag{A.2}$$

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

The proof involves the following steps:

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

Points 2. and 3. imply by Schauder's fixed-point theorem that Z has a fixed point, which by point 1. yields the desired result. The formal statement of Schauder's 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) satisfies rational expectations and are market clearing note that $Z^\lambda(p,\lambda) = \lambda$ implies that $\lambda = \lambda(A(p,\lambda))$. Therefore, (p,λ) satisfies rational expectations. $Z^d(p,\lambda) = p$ implies $D^c(p,\lambda) \leq q^c$ for all $c \in C$. Moreover, for any $c \in C$, if $D^c(p,\lambda) < q^c$ then it must be that $p^c = 0$. Therefore, (p,λ) is market clearing.

2. It is clear that K is nonempty.

   To show that K is convex, 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.

Lemma A.1. *Λ is convex.*

Proof. Take any two functions $\psi, \hat{\psi} \in \Lambda$ and any $\beta \in [0,1]$. We must show that the function $\tilde{\psi} := \beta\psi + (1-\beta)\hat{\psi}$ is in Λ. To see that this is the case note that $\tilde{\psi}^x \in [0,1]$ for any $x \in [0,1]$ and any $\beta \in [0,1]$, as $\psi^x, \hat{\psi}^x \in [0,1]$ and $\tilde{\psi}^x \in [\min\{\psi^x, \hat{\psi}^x\}, \max\{\psi^x, \hat{\psi}^x\}]$. $\tilde{\psi}$ must be also be non-decreasing; for any $x < y$ with $x, y \in [0,1]^{N+1}$, $\tilde{\psi}^x = \beta\psi^x + (1-\beta)\hat{\psi}^x \leq \beta\psi^y + (1-\beta)\hat{\psi}^y = \tilde{\psi}^y$ where the inequality follows from the non-decreasing property of ψ and $\hat{\psi}$. As $\tilde{\psi}$ is a non-decreasing function from $[0,1]^{N+1}$ to $[0,1]$, $\tilde{\psi} \in \Lambda$, i.e. Λ is convex. □

To show that K is compact and Hausdorff, we note that $[0,1]$ is clearly compact and Hausdorff. It remains to show that Λ is compact and Hausdorff, which then implies that K is compact and Hausdorff as the product of compact and Hausdorff sets.

Lemma A.2. *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 $\langle \psi_\ell \rangle_{\ell=1,2,\ldots}$ be a convergent Moore-Smith sequence with limit ψ, where each $\psi_\ell \in \Lambda$. We need to show that $\psi \in \Lambda$. For any $x < y$ with $x, y \in [0,1]^{N+1}$ and any ℓ it must be the case that $0 \leq \psi_\ell^x \leq \psi_\ell^y \leq 1$. Taking the limit with respect to ℓ yields that $0 \leq \psi^x \leq \psi^y \leq 1$, i.e. $\psi \in \Lambda$. Therefore, Λ is compact. Similarly Λ is Hausdorff: $\Lambda \subset [0,1]^{[0,1]}$ is Hausdorff as a subset of a Hausdorff space. □

3. Consider any pairs $(p,\lambda) \in [0,1]^{N+1} \times \Lambda^{N+1}$ and $(p',\lambda') \in [0,1]^{N+1} \times \Lambda^{N+1}$ where we write $\alpha = A(p,\lambda)$ and $\alpha' = A(p',\lambda')$. We must show that for any $\epsilon > 0$ there exists $\delta > 0$ such that if $||(p,\lambda) - (p',\lambda')||_\infty < \delta$ then $||Z(p,\lambda) - Z(p',\lambda')||_\infty < \epsilon$. Note that by construction, $Z^{d,c}(p,\lambda)$ is continuous in $D^c(p,\lambda)$ for all $c \in C$ (this follows from Equation A.1 and noting that $D^c(\cdot,\cdot) \leq 1 < 1 + q^c$). Also, $Z^\lambda(p,\lambda) = \lambda(\alpha)$ and $Z^\lambda(p',\lambda') = \lambda(\alpha')$ by Equation A.2. Therefore, it suffices to show that for any $\epsilon > 0$ there exists $\delta > 0$ such that if $||(p,\lambda) - (p',\lambda')||_\infty < \delta$ then both $|D^c(p,\lambda) - D^c(p',\lambda')| < \epsilon$ for all $c \in C$ and $||\lambda(\alpha) - \lambda(\alpha')||_\infty < \epsilon$.

   By Assumption **A1**, for almost all $\theta \in \Theta$ we have that $\alpha(\theta) \neq \alpha'(\theta)$ if and only if $D^\theta(p,\lambda) \neq D^\theta(p',\lambda')$. Denote the set of students for whom $D^\theta(p,\lambda) \neq D^\theta(p',\lambda')$ as $\Theta(\alpha,\alpha') := \{\theta | D^\theta(p,\lambda) \neq D^\theta(p',\lambda')\}$. We first argue that for sufficiently small δ, $\eta(\Theta(\alpha,\alpha')) < \epsilon$. Note that $\theta \in \Theta(\alpha,\alpha')$ only if at least one of the following holds:

   (a) $\{c | p^c \leq r^{\theta,c}\} \neq \{c | p'^c \leq r^{\theta,c}\}$ (different choice sets), or

   (b) $\succeq^{\theta|\lambda} \neq \succeq^{\theta|\lambda'}$ (different ordinal rankings).

Let the set of students with different choice sets be denoted $\Theta^1(\alpha,\alpha') := \{\theta | \{c | p^c \leq r^{\theta,c}\} \neq \{c | p'^c \leq r^{\theta,c}\}\}$, and the set of students with different ordinal preferences $\Theta^2(\alpha,\alpha') := \{\theta | \succeq^{\theta|\lambda} \neq \succeq^{\theta|\lambda'}\}$.

For any $\delta < 1$, when $||(p,\lambda)-(p',\lambda')||_\infty < \delta$ the measure of students with different choice sets $\eta(\Theta^1(\alpha,\alpha')) < (N+1)\delta$ by construction. This is due to the fact that $|p^c - p'^c| < \delta$ for all programs $c \in C$ and the ongoing assumption of a uniform distribution of student scores within program. Let $\epsilon' = \frac{\epsilon}{N+2}$. By Assumption **A4**, there exists $\delta^1 > 0$ such that when $||(p,\lambda)-(p',\lambda')||_\infty < \delta^1$ the measure of students with different ordinal rankings $\eta(\Theta^2(\alpha,\alpha')) < \epsilon'$. Let $\delta = \min\{\frac{\epsilon}{N+2}, \delta^1\}$. Therefore, if $||(p,\lambda)-(p',\lambda')||_\infty < \delta$ it must be the case that

$$\begin{aligned}
\eta(\Theta(\alpha,\alpha')) &\leq \eta(\Theta^1(\alpha,\alpha') \cup \Theta^2(\alpha,\alpha')) \\
&\leq \eta(\Theta^1(\alpha,\alpha')) + \eta(\Theta^2(\alpha,\alpha')) \\
&< (N+1)\delta + \epsilon' \\
&\leq (N+1)\frac{\epsilon}{N+2} + \frac{\epsilon}{N+2} \\
&= \epsilon
\end{aligned} \tag{A.3}$$

where the first inequality holds because $\theta \in \Theta(\alpha,\alpha')$ only if θ is an element of at least one of $\Theta^1(\alpha,\alpha')$ and $\Theta^2(\alpha,\alpha')$. Therefore, the proof is complete if we can show that

$$\eta(\Theta(\alpha,\alpha')) \geq |D^c(p,\lambda) - D^c(p',\lambda')| \text{ for all } c \in C \tag{A.4}$$

and

$$\eta(\Theta(\alpha,\alpha')) \geq ||\lambda(\alpha) - \lambda(\alpha)||_\infty \tag{A.5}$$

To see that Inequality A.4 holds, note that for any $c \in C$ we have that

$$\begin{aligned}
\eta(\Theta(\alpha,\alpha')) &= \frac{1}{2}\sum_{c \in C}\left[\eta(\alpha(c) \setminus \alpha'(c)) + \eta(\alpha'(c) \setminus \alpha(c)\right] \\
&\geq \eta(\alpha(c) \setminus \alpha'(c)) + \eta(\alpha'(c) \setminus \alpha(c)) \\
&= \eta(\alpha(c)) + \eta(\alpha'(c)) - 2\eta(\alpha(c) \cap \alpha'(c)) \\
&= \max\{\eta(\alpha(c)), \eta(\alpha'(c))\} + \min\{\eta(\alpha(c)), \eta(\alpha'(c))\} - 2\eta(\alpha(c) \cap \alpha'(c)) \\
&\geq \max\{\eta(\alpha(c)), \eta(\alpha'(c))\} - \min\{\eta(\alpha(c)), \eta(\alpha'(c))\} \\
&= |\eta(\alpha(c)) - \eta(\alpha'(c))| \\
&= |D^c(p,\lambda) - D^c(p',\lambda')|
\end{aligned} \tag{A.6}$$

The first equality follows because each student $\theta \in \Theta(\alpha,\alpha')$ is double counted in the RHS of the top line.[3] The first inequality follows because the total measure of students with different assignments with respect to α and α' is weakly greater than the measure of students who are assigned to program c in exactly one of the two assignments. The second inequality follows because $\min\{\eta(\alpha(c)),\eta(\alpha'(c))\} \geq \eta(\alpha(c) \cap \alpha'(c))$.

To see that Inequality A.5 holds, note that for any $c \in C$ and any $x \in [0,1]^{N+1}$,

$$\begin{aligned}\eta(\Theta(\alpha,\alpha')) &\geq |\eta(\alpha(c)) - \eta(\alpha'(c))| \\ &\geq |\lambda^{c,x}(\alpha) - \lambda^{c,x}(\alpha')| \\ &\geq ||\lambda(\alpha) - \lambda(\alpha')||_\infty\end{aligned}$$

where the first inequality follows from Inequalities A.6, the second inequality follows because the difference in the measure of students with scores below x assigned to c at α and α' cannot be larger than the total measure of students who are assigned to c in only one of α and α', and the final inequality follows from the definition of the $||\cdot||_\infty$ norm.

This completes the proof of continuity, and therefore, the proof of the theorem.

□

Remark 1

Proof. We show this result via the following example:

Example 3. *Let $C = \{c_0, c_1, c_2\}$ and let each program have common rankings over student types, that is $r^\theta := r^{\theta,c_0} = r^{\theta,c_1} = r^{\theta,c_2}$. Programs c_1 and c_2 have identical capacities $q^{c_1} = q^{c_2} < \frac{1}{2}$. For $i \in \{1,2\}$ and any assignment α let*

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

that is, $s^{c_i}(\lambda(\alpha))$ is the mean score of students assigned to c_i with respect to α.

For any α, all students prefer to be assigned to either c_1 or c_2 to being unassigned. Therefore, we omit descriptions of the students assigned to c_0 in what follows. 2ϵ measure of students 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 α and an ϵ measure of students who prefer c_2 to c_1 for any α, 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 $s^{c_i}(\lambda(\alpha)) - s^{c_j}(\lambda(\alpha)) > \frac{q}{2}$ and $\lambda^{c_i,(1,1,1)}(\alpha) > \epsilon$

[3] That is, if $\theta \in \alpha(c_1) \cap \alpha'(c_2)$ then θ contributes to the sum on the RHS for both c_1 and c_2.

for each $i \in \{1,2\}$. This example is consistent with our regularity conditions.[4]

Let

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

and

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

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. Given our assumption that $\epsilon \leq \frac{1}{2}$, $p^{c_i} \leq p^{c_j}$. Moreover, as $\epsilon \to 0$, $s^{c_i} - s^{c_j} \to q > \frac{q}{2}$. Therefore, for sufficiently small ϵ, (p, λ) is market clearing because all students with scores weakly above p^{c_i} (except for those who have weak peer preferences and prefer c_j) prefer to attend c_i and all remaining students with scores weakly above p^{c_j} prefer to attend c_j to remaining unmatched. This also implies that (p, λ) satisfies rational expectations for sufficiently small ϵ. 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 $\epsilon < \epsilon^*$.

□

Proposition 1

Proof.

1. Let μ_* be a stable matching. As we argue in Remark 2, letting $\succ$ represent a profile of ROLs such that $\succ^\theta = \succeq^{\theta|\mu_*}$ for all θ, $\varphi(\succ) = \mu_*$ for any stable mechanism φ. For each θ, let $\succ^\theta$ be the submitted preferences for θ such that $\mu_*(\theta)$ is the unique acceptable program, and let $\succ$ be the profile of such reports for all $\theta \in \Theta$. Because φ is stable, $\varphi(\succ) = \varphi(\succ) = \mu_*$. To see that this is a Bayes Nash equilibrium, note that for any θ and any program $c \succ^{\theta|\mu_*} \mu_*(\theta)$, stability of μ_* implies that there is no deviating report $\succ'^\theta \neq \succeq^\theta$ that will result in θ matching with c.

   Suppose for contradiction that $\succ$ is a Bayes Nash equilibrium of φ but that $\mu = \varphi(\succ)$ 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 $\succ$. Let p be the associated cutoff vector. Now consider reported preferences $\hat{\succ}$ where $\hat{\succ}^{\theta'} = \succ^{\theta'}$ 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

[4] Note that we have only specified student ordinal preferences under certain conditions, meaning there are many utility functions that satisfy our regularity assumptions and comport with this example.

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 Bayes 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 $p^c \leq \max\{1-q^c, 0\} < 1$ for all $c \in C$. Let $\tilde{p}$ be an $N+1$ dimensional vector such that $p < \tilde{p} < 1$.

   Consider any student θ such that $r^\theta \geq \tilde{p}$. By Assumption **A1**, $\succeq^{\theta|\mu_*}$ is strict for almost all such θ, and we proceed assuming $\succeq^{\theta|\mu_*}$ is strict. By the stability of μ_* and the fact that θ's score $r^{\theta,c}$ at each program c exceeds c's cutoff, it must be the case that $\mu_*(\theta)$ is the $\succeq^{\theta|\mu_*}$-maximal program.

   Moreover, it follows from Assumption **A2** that for each program $c \in C$ there exists some student θ^c such that $p < r^{\theta^c} < r^\theta$ and such that c is the unique $\succeq^{\theta^c|\mu_*}$-maximal program. By the stability hypothesis, $\mu_*(\theta^c) = c$. Because φ respects rankings, this implies that θ is admitted to her top-ranked program according to her submitted preferences $\tilde{\succ}^\theta$. Therefore, stability implies that $\mu_*(\theta)$ is θ's top-ranked program according to $\tilde{\succ}^\theta$. By the equilibrium hypothesis, it must be that the $\tilde{\succ}^\theta$-maximal program is the same as the $\succeq^{\theta|\mu(\sigma^\theta,\tilde{\succ})}$-maximal program. That is, it must be that, for equilibrium profile $\tilde{\succ}$, student θ realizes that she will receive her top-ranked program, and therefore, her top-ranked program must coincide with the top-ranked program according to her true preferences (given her beliefs over the distribution of types).

   The logic of the previous two paragraphs implies that the top-ranked program according to $\succeq^{\theta|\mu_*}$ coincides with the top-ranked program according to $\succeq^{\theta|\mu(\sigma^\theta,\tilde{\succ})}$ for almost all θ with $r^\theta \geq \tilde{p}$. But this contradicts the ongoing assumption that $\eta(L_{\tilde{\succ},\varphi,\tilde{p}}) > 0$.

□

Before proceeding, we provide a lemma which is useful in several upcoming proofs.

Lemma A.3. *Let E be a market satisfying **A4**. For any $\lambda, \lambda' \in \Lambda^{N+1}$ we write $p = P\lambda$, $p' = P\lambda'$, $\mu = A(P\lambda,\lambda)$, and $\mu' = A(P\lambda',\lambda')$. For any $\epsilon > 0$ there exists $\delta > 0$ such that if $||\lambda - \lambda'||_\infty < \delta$ then $||p - p'||_\infty < \epsilon$ and $\eta\{\theta|\mu(\theta) \neq \mu'(\theta)\} < \epsilon$.*

Proof. Fix $\epsilon > 0$, and let ω define the bound on the support of student types in Assumption **A2**. We first argue that there exists $\delta > 0$ such that $||p - p'||_\infty < \epsilon$ when $||\lambda - \lambda'||_\infty < \delta$. If $p = p'$ then we are done. In the complementary case, assume without loss of generality that $p^c > p'^c$ for some $c \in C$.

By Assumption **A4** there exists $\delta > 0$ such that $\eta(\{\theta|\succeq^{\theta|\lambda} \neq \succeq^{\theta|\lambda'}\}) < \epsilon\omega$ when $||\lambda - \lambda'||_\infty < \delta$. Then

for such λ,λ'

$$\begin{aligned}\epsilon\omega &> \eta(\{\theta|\succeq^{\theta|\lambda}\neq\succeq^{\theta|\lambda'}\})\\ &\geq D^c(p,\lambda)-D^c(p,\lambda')\\ &=q^c-D^c(p,\lambda')\\ &\geq 0\end{aligned} \tag{A.7}$$

where the second inequality follows from Inequalities A.6 in the case where $\alpha=A(p,\lambda)$ and $\alpha'=A(p,\lambda')$ (i.e. given the same admissions cutoffs, a weakly larger measure of students have different ordinal preferences than different demands) and the ongoing assumption that $p^c>p'^c$, the equality follows because the assumption that $p^c>p'^c$ implies that $p^c>0$ which therefore implies that $D^c(p,\lambda)=\eta(\mu(c))=q^c$.

In order to respect c's capacity constraint, Inequality A.7 implies that there is at most a $\epsilon\omega$ measure of students matched to c in μ' with scores below p^c, $\eta\{\theta\in\mu'(c)|r^{\theta,c}<p^c\}\leq\epsilon\omega$. By bound ω from Assumption **A2**, it must be that $p'^c\in(p^c-\epsilon,p^c)$ when $||\lambda-\lambda'||_\infty<\delta$. Applying this argument across all programs $c\in C$ implies that $||p-p'||_\infty<\epsilon$ when $||\lambda-\lambda'||_\infty<\delta$.

That $\eta\{\theta|\mu(\theta)\neq\mu'(\theta)\}<\epsilon$ follows from the above argument and Inequality A.3. □

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 $(P\lambda_*,\lambda_*)$ satisfies 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. **"Only if" part** Fix any $\epsilon>0$. We want to show that there exists $\delta>0$ such that if $||\lambda_t-\lambda_{t-1}||_\infty<\delta$ then μ_t is $\epsilon-$stable.

   Let B denote the set of students who block μ_t, that is $B:=\{\theta|(\theta,c)$ block μ_t for some $c\in C\}$. Let $B^{\lambda,\lambda'}:=\{\theta|D^\theta(p_t,\lambda_t)\neq D^\theta(p_t,\lambda_{t-1})\}$. The following result states that almost surely $\theta\in B$ if and only if $\theta\in B^{\lambda,\lambda'}$.

   Lemma A.4. $\eta(\{B\setminus B^{\lambda,\lambda'}\}\cup\{B^{\lambda,\lambda'}\setminus B\})=0$.

   Proof. We prove this result by showing that $\eta(B\setminus B^{\lambda,\lambda'})=0$ and $\eta(B^{\lambda,\lambda'}\setminus B)=0$. This implies that $\eta(\{B\setminus B^{\lambda,\lambda'}\}\cup\{B^{\lambda,\lambda'}\setminus B\})\leq\eta(B\setminus B^{\lambda,\lambda'})+\eta(B^{\lambda,\lambda'}\setminus B)=0$.

   For each $\theta\in B$ there exists some $c^\theta\in C$ such that (θ,c^θ) block μ_t. Because $\mu_t\in M$, there is an associated cutoff vector p_t such that $r^{\theta,c^\theta}\geq p^c$ and $c^\theta\succ^{\theta|\mu_t}\mu_t(\theta)$, which implies that $D^\theta(p_t,\lambda_t)\neq D^\theta(p_t,\lambda_{t-1})$ Therefore, $\eta(B\setminus B^{\lambda,\lambda'})=0$.

By Assumption **A1**, for almost all $\theta \in B^{\lambda,\lambda'}$ there exists a unique $c^\theta = D^\theta(p_t,\lambda_t)$. If $c^\theta \neq D^\theta(p_t,\lambda_{t-1})$ then (θ,c^θ) form a blocking pair at μ_t for almost all $\theta \in B^{\lambda,\lambda'}$. Therefore, $\eta(B^{\lambda,\lambda'} \setminus B) = 0$. □

Returning to the proof of the theorem, by Assumption **A4** there exists $\delta > 0$ such that if $\|\lambda_{t-1} - \lambda_t\|_\infty < \delta$, then $\eta(\{\theta | \succeq^{\theta|\mu_{t-1}} \neq \succeq^{\theta|\mu_t}\}) < \epsilon$. For almost all $\theta \in B^{\lambda,\lambda'}$ it is the case that $\succeq^{\theta|\mu_t - 1} \neq \succeq^{\theta|\mu_t}$, i.e. some subset of students whose ordinal rankings change demand a different program given p. Therefore, if $\|\lambda_{t-1} - \lambda_t\|_\infty < \delta$, similar logic as in Inequality A.7 implies

$$\begin{aligned}\eta(B) &= \eta(B^{\lambda,\lambda'}) \\ &\leq \eta(\{\theta | \succeq^{\theta|\mu_{t-1}} \neq \succeq^{\theta|\mu_t}\}) \\ &< \epsilon\end{aligned}$$

where the equality follows from Lemma A.4. Therefore, for $\|\lambda_{t-1} - \lambda_t\|_\infty < \delta$, $\eta(B) < \epsilon$ as desired.

"If" part Fix any $\delta > 0$ and let B be the set of students involved in at least one blocking pair at μ_t. We wish to show that there exists $\epsilon > 0$ such that if $\eta(B) < \epsilon$ then $\|\lambda_{t-1} - \lambda_t\|_\infty < \delta$.

Consider three alternative markets $E_t = [\zeta^{\eta,\mu_{t-1}}, q, N]$, $E_{t+1} = [\zeta^{\eta,\mu_t}, q, N]$, and $E_\gamma = [\zeta^\gamma, q, N]$. We define measure ζ^γ as follows for $\gamma \in (0,1)$:

- For all $\theta \in \Theta$ and any assignment α, $\succeq^{\theta|\alpha} \in \{\succeq^{\theta|\mu_{t-1}}, \succeq^{\theta|\mu_t}\}$,
- for any open set $R \subset [0,1]^{N+1}$, any assignment α, and any $\succeq \in P$, $\zeta^\gamma(\{\theta | r^\theta \in R \cap B \text{ and } \succeq^{\theta|\alpha} = \succeq\}) = \eta(\{\theta | r^\theta \in R \cap B \text{ and } \succeq^{\theta|\mu_{t-1}} = \succeq\})$, and
- for any open set $R \subset [0,1]^{N+1}$, any assignment α, and any $\succeq \in P$, $\zeta^\gamma(\{\theta | r^\theta \in R \cap \{\Theta \setminus B\} \text{ and } \succeq^{\theta|\alpha} = \succeq\}) = (1-\gamma)\eta(\{\theta | r^\theta \in R \cap \{\Theta \setminus B\} \text{ and } \succeq^{\theta|\mu_t} = \succeq\})$.

That is, ζ^γ specifies student types such that students involved in blocking pairs at μ_t in market E have the same preferences as in market E_t and $1-\gamma$ fraction of students selected "uniformly at random" among those not involved in blocking pairs have the same preferences as in market E_{t+1}. Let μ_t and μ_{t+1}, be the (unique) stable matchings in E_t and E_{t+1}. Recall that by Remark 2, μ_t and μ_{t+1} are the outcomes of the TIM process at times t and $t+1$, respectively.

We proceed with the proof first by showing that there is a unique stable matching μ_γ in market E_γ, and that $\mu_t = \mu_\gamma$ for all $\gamma \in (0,1)$. Then we show that for sufficiently small γ and ϵ, $\mu_\gamma(\theta) = \mu_{t+1}(\theta)$ for at least $1-\delta$ measure of students. This implies that $\mu_t(\theta) = \mu_{t+1}(\theta)$ for at least $1-\delta$ measure of students, completing the proof.

Lemma A.5. *For any $\gamma \in (0,1)$ there is a unique stable matching $\mu_\gamma = \mu_t$ in market E_γ.*

Proof. To see that there is a unique stable matching μ_γ in market E_γ, note that E_γ satisfies Assumption **A2**: letting ω_t and ω_{t-1} be a bound on the support of ordinal preferences in markets E_t and E_{t-1}, respectively, by construction, $\min\{\omega_{t-1},(1-\gamma)\omega_t\}$ satisfies **A2** in E_γ. By Theorem 1 of Azevedo and Leshno (2016), there is therefore a unique stable matching μ_γ in E_γ.

To see that $\mu_\gamma(\theta) = \mu_t(\theta)$ for almost all θ and for all $\gamma \in (0,1)$, we claim that μ_t is stable in E_γ. By Lemma A.4, θ blocks matching μ_t in market E (i.e. $\theta \in B$) only if $D^\theta(p_t,\lambda_t) \neq D^\theta(p_t,\lambda_{t-1})$ (excepting a measure zero set of students for which $D^\theta(p_t,\lambda_t)$ or $D^\theta(p_t,\lambda_{t-1})$ is not a singleton, by Assumption **A1**). By construction, there is no positive measure set of students for which $D^\theta(p_t,\lambda_t) \neq D^\theta(p_t,\lambda_{t-1})$ in market E_γ. Therefore, μ_t is stable in E_γ, and by the earlier argument, is the unique stable matching in E_γ. □

Lemma A.6. *There exists $\gamma^* > 0$ and $\epsilon^* > 0$ such that if $\gamma < \gamma^*$ and $\epsilon < \epsilon^*$ then $\eta(\{\theta|\mu_\gamma(\theta) = \mu_{t+1}(\theta)\} < \delta$.*

Proof. Follows directly from Lemma A.3. □

The previous two Lemmas establish that there exists ϵ such that if $\eta(B) < \epsilon$ then $\eta(\{\theta|\mu_\gamma(\theta) = \mu_{t+1}(\theta)\} < \delta$ which in turn implies that $\|\lambda_{t-1} - \lambda_t\|_\infty < \delta$ as desired.

□

Remark 3

Proof. Consider any market $E = [\eta,q,N]$ such that $\eta \in \hat{H}(N)$, and let $f \mapsto \eta$. Consider any ability distribution λ. By Assumption **A1** almost all students have strict preferences induced by λ, that is, for any two programs c,c', $c \succeq^{\theta|\lambda} c'$ and $c' \succeq^{\theta|\lambda} c$ for almost no students. Fix any $\epsilon > 0$. By the uniform continuity of $f^{\theta,c}$ for almost all θ and all $c \in C \setminus \{c_0\}$, there exists some $\delta > 0$ such that for any ability distribution λ' with $||\lambda - \lambda'||_\infty < \delta$ we have that $\eta(\{\theta| \succeq^{\theta|\lambda} = \succeq^{\theta|\lambda'}\}) > 1-\epsilon$. Then E satisfies **A4**: $\eta(\{\theta| \succeq^{\theta|\lambda} \neq \succeq^{\theta|\lambda'}) < \epsilon$ for any λ' with $||\lambda - \lambda'||_\infty < \delta$. By Theorem 1, E has at least one stable matching.

□

Theorem 3

Proof.

1. We show the desired result by constructing a measure $\hat{\eta}$ by "scaling up" the intrinsic component of student preferences such that the difference between the cardinal values for any two programs is greater than the range that function f can take. Formally, suppose that for almost all students θ and all distinct $c,c' \in C$, either $v^{\theta,c} - v^{\theta,c'} > b - a$ or $v^{\theta,c'} - v^{\theta,c} > b - a$. Therefore, $u^\theta(c|\alpha) > u^\theta(c'|\alpha)$ for all α if and only if $v^{\theta,c} > v^{\theta,c'}$. As a result, almost all students will have the same ordinal preferences over programs given μ_0 and μ_1, implying that $D^\theta(p_1,\lambda_1) = D^\theta(p_2,\lambda_2)$ for almost all students. Therefore, the TIM process reaches steady state at $t = 1$.

2. We first argue that the set of measures that admit a negative externality group is open in $\hat{H}(N)$ with respect to the $||\cdot||_p$ norm. To see this, suppose that $\eta \in \hat{H}(N)$, let $F \mapsto \eta$ be a canonical representation, and suppose that at program $c' \in C \setminus \{c_0\}$ and assignment α, $\Theta^I \subset \alpha(c')$ and $\Theta^O \subset \Theta \setminus \alpha(c')$ form a negative externality group. It suffices to show that for any $\epsilon > 0$ and any family of functions $\{\tilde{F}^{\theta,c}\}_{c \in C, \theta \in \Theta}$ with $||\tilde{F}^{\theta,c} - F^{\theta,c}||_\infty < \epsilon$, there exist a program $\tilde{c} \in C \setminus \{c_0\}$, an assignment $\tilde{\alpha}$, $\tilde{\Theta}^I \subset \tilde{\alpha}(\tilde{c})$ and $\tilde{\Theta}^O \subset \Theta \setminus \tilde{\alpha}(\tilde{c})$ with $\eta(\tilde{\Theta}^I) \geq \eta(\tilde{\Theta}^O)$ such that $\tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha})) < F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I))$ for all $\theta \in \tilde{\Theta}^I$.

   By the uniform continuity of $F^{\theta,c}(\cdot)$ for almost all θ,c, it must be the case that there exists some $\delta > 0$ such that some subset $\tilde{\Theta}^I \subset \Theta^I$ with $\eta(\tilde{\Theta}^I) > \eta(\Theta^I) - \delta$, $\Theta^I \subset \alpha(c')$ and Θ^O form a negative externality group at program $c' \in C \setminus \{c_0\}$ and assignment α. Toward the desired construction take $\epsilon = \delta$, and let $\tilde{c} = c'$, $\tilde{\alpha}(\tilde{c}) = \alpha(c')$, $\tilde{\Theta}^O = \Theta^O$, and let $\tilde{\Theta}^I$ be defined as above. Then for any $\{\tilde{F}^{\theta,c}\}_{\theta \in \Theta, c \in C}$ with $||\tilde{F}^{\theta,c} - F^{\theta,c}||_\infty < \epsilon = \delta$, $\tilde{\Theta}^I$ and $\tilde{\Theta}^O$ form a negative externality group at program $\tilde{c}$ and assignment $\tilde{\alpha}$.

   We now argue that the set of measures that admit a negative externality group is dense in $\hat{H}(N)$ with respect to the $||\cdot||_p$ norm. Fix $\eta \in \hat{H}(N)$ and let $F \mapsto \eta$ be a canonical representation. It suffices to show that for any ϵ there exists a distinct family of functions $\{\tilde{F}^{\theta,c}\}_{\theta \in \Theta, c \in C}$ with $||\tilde{F}^{\theta,c} - F^{\theta,c}||_\infty < \epsilon$ where $\tilde{F} \mapsto \tilde{\eta}$, a program $\tilde{c} \in C \setminus \{c_0\}$, an assignment $\tilde{\alpha}$, and sets $\tilde{\Theta}^I \subset \tilde{\alpha}(\tilde{c})$ and $\tilde{\Theta}^O \subset \Theta \setminus \tilde{\alpha}(\tilde{c})$ such that $\tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha})) < \tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I))$ for all $\theta \in \tilde{\Theta}_I$.

   Fix any assignment $\tilde{\alpha}$ such that there exists a program $\tilde{c} \in C \setminus \{c_0\}$ where $\tilde{\eta}(\tilde{\alpha}(\tilde{c})) > 0$. By the uniform continuity of $F^{\theta,c}(\cdot)$, it must be the case that there exists some $\delta > 0$ such that for any $\tilde{\Theta}^I \subset \tilde{\alpha}(\tilde{c})$ with $\tilde{\eta}(\tilde{\Theta}^I) < \delta$, $|F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha}(\tilde{c}))) - F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I))| < \frac{\epsilon}{2}$ for all $\theta \in \Theta$. Let $\tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha})) = \min\{F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha})), F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I))\}$, and $\tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I)) = \max\{F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha})), F^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I))\}$. Therefore, $\{\tilde{F}^{\theta,c}\}_{\theta \in \Theta, c \in C}$ admits a negative externality group.

   Note that $\{\tilde{F}^{\theta,c}\}_{c \in C, \theta \in \Theta}$ is a canonical representation, since the "min," "max" construction preserves the renormalization construction. Also note that for δ sufficiently small, the full support assumption **A2** cannot be violated.

   Therefore, all that remains to be shown is that the remainder of the family of functions $\{\tilde{F}^{\theta,c}\}_{\theta \in \Theta, c \in C}$ can be constructed to satisfy uniform continuity. Based on fact that $F^{\theta,c}(\cdot)$ is uniformly continuous, and the construction of ϵ, $\tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\alpha}))$, and $\tilde{F}^{\theta,\tilde{c}}(\lambda^{\tilde{c}}(\tilde{\Theta}^O \cup \tilde{\alpha}(\tilde{c}) \setminus \tilde{\Theta}^I))$, such a family of functions clearly exists for all $\theta \in \tilde{\Theta}^I$ and $\tilde{c}$. We can additionally define $\tilde{F}^{\theta,c}(\cdot) = F^{\theta,c}(\cdot)$ for all (θ,c) such that either $\theta \notin \Theta^I$ or $c \neq \tilde{c}$. Therefore, there exists $\tilde{F}^{\theta,c}(\cdot) = F^{\theta,c}(\cdot)$ that both satisfies uniform continuity and admits a negative externality group.

3. We prove this result for $N = 1$ and discuss how it easily extends to the case in which $N > 1$.

   Let $\eta \in H(1) \setminus \hat{H}(1)$. Then there exist an assignment $\hat{\alpha}$ and positive measure sets of students

$\Theta^I \subset \hat{\alpha}(c_1)$ and $\Theta^O \subset \Theta\hat{\alpha}(c_1)$ where $\eta(\Theta^I) \geq \eta(\Theta^O)$ such that $f^{\theta,c_1}(\lambda^{c_1}(\Theta^O \cup \hat{\alpha}(c_1) \setminus \Theta^I) > f^{\theta,c_1}(\lambda^{c_1}(\hat{\alpha}(c_1))$ for all $\theta \in \Theta^I$.

Fix $\epsilon > 0$ and let $\omega > 0$ be such that for almost all θ, $|f^{\theta,c_1}(\lambda_1^c) - f^{\theta,c_1}(\hat{\lambda}_1^c)| < \epsilon$ if $\lambda^{c_1}, \hat{\lambda}^{c_1} \in \Lambda$ are such that $||\lambda^{c_1} - \hat{\lambda}^{c_1}||_\infty < \omega$. Let ω be the relevant lower bound on the support of student types, following Assumption **A2**.

We construct the desired market $E = [\hat{\eta}, q, 1]$ as follows:

- Let there be the following five disjoint sets of students: Ω such that $\hat{\eta}(\Omega) = \omega$, $\hat{\Theta}^I \subset \Theta^I$ where $\hat{\eta}(\hat{\Theta}^I) = (1-\omega)\hat{\eta}(\Theta^I)$, $\hat{\Theta}^O \subset \Theta^O$ where $\hat{\eta}(\hat{\Theta}^O) = (1-\omega)\hat{\eta}(\Theta^O)$, $\hat{\Theta}^U \subset \hat{\alpha}(c_1) \setminus \Theta^I$ where $\hat{\eta}(\hat{\Theta}^U) = (1-\omega)\hat{\eta}(\hat{\alpha}(c_1) \setminus \Theta^I)$, and $\hat{\Theta}^L$ where $\hat{\eta}(\hat{\Theta}^L) = (1-\omega)[1 - \hat{\eta}(\hat{\Theta}^O) - \hat{\eta}(\hat{\Theta}^I) - \hat{\eta}(\hat{\Theta}^L) - \hat{\eta}(\hat{\Theta}^U)]$. It is easy to see that $\hat{\eta}(\Omega \cup \hat{\Theta}^I \cup \hat{\Theta}^L \cup \hat{\Theta}^O \cup \hat{\Theta}^U) = 1$.
- $\hat{u}^\theta(c_1|\alpha) = \hat{v}^{\theta,c} + f^{\theta,c_1}(\lambda^{c_1}(\alpha))$ for each $\theta \in \Theta$ such that:
  - $\hat{u}^\theta(c_1|\hat{\alpha}) + \epsilon < \hat{u}^\theta(c_0|\hat{\alpha})$ for all $\theta \in \hat{\Theta}^I$,
  - $\hat{u}^\theta(c_1|\alpha) > \hat{u}^\theta(c_0|\alpha)$ for all α and for all $\theta \in \hat{\Theta}^0$,
  - $\hat{u}^\theta(c_1|\alpha) > \hat{u}^\theta(c_0|\alpha)$ for all α and for all $\theta \in \Theta^U$,
  - $\hat{u}^\theta(c_1|\alpha) > \hat{u}^\theta(c_0|\alpha)$ for all α and for all $\theta \in \Omega$, and
  - $\hat{u}^\theta(c_0|\alpha) > \hat{u}^\theta(c_1|\alpha)$ for all α and for all $\theta \in \Theta^L$.
- Scores are defined implicitly by the following:
  - For any $x \in [0,1]$, $\hat{\eta}(\theta \in \Omega | r^{\theta,c_1} < x) = \omega x$,
  - $r^{\theta^L,c_1} < r^{\theta^O,c_1} < r^{\theta^I,c_1} < r^{\theta^U,c_1}$ for any $\theta^L \in \hat{\Theta}^L$, any $\theta^O \in \hat{\Theta}^O$, any $\theta^I \in \hat{\Theta}^I$, and any $\theta^U \in \hat{\Theta}^U$.
- $q^{c_1} = (1+\omega)\hat{\eta}(\hat{\Theta}^I \cup \hat{\Theta}^U)$.

Let $\mu_0(c_1) = \hat{\Theta}^I \cup \hat{\Theta}^U \cup \{\theta \in \Omega | r^{\theta,c_1} \geq 1 - (1+\omega)\hat{\eta}(\hat{\Theta}^I \cup \hat{\Theta}^U)\}$. Therefore, by construction of q, $\hat{\eta}(\mu_0(c_1)) = q^{c_1}$. Note that $||\lambda^{c_1}(\mu_0(c_1)) - \lambda^{c_1}(\hat{\alpha}(c_1))||_\infty < \omega$, and so by the construction of preferences and the uniform continuity of $f^{\theta,c_1}(\cdot)$, all students $\theta \in \hat{\Theta}^U \cup \hat{\Theta}^O \cup \Omega$ have preferences $\hat{u}^\theta(c_1|\mu_0) > u^\theta(c_0|\mu_0)$, and all student $\theta \in \hat{\Theta}^I \cup \hat{\Theta}^L$ have preferences $\hat{u}^\theta(c_0|\mu_0) > u^\theta(c_1|\mu_0)$. Given these preferences and the student scores defined above, $\mu_1(c_1) = \hat{\Theta}^U \cup \hat{\Theta}^O \cup \{\theta \in \Omega | r^{\theta,c_1} \geq \tau\}$, where τ is defined implicitly by the infimum value of $x \geq 0$ such that $\hat{\eta}(\hat{\Theta}^U) + \hat{\eta}(\hat{\Theta}^O) + \hat{\eta}(\{\theta \in \Omega | r^{\theta,c_1} \geq x\}) < q^{c_1}$. Note that $||\lambda^{c_1}(\mu_1(c_1)) - \lambda^{c_1}(\Theta^O \cup \hat{\alpha}(c_1) \setminus \Theta^I)||_\infty < \omega$, and so by the construction of preferences and the uniform continuity of $f^{\theta,c_1}(\cdot)$, all student $\theta \in \hat{\Theta}^U \cup \hat{\Theta}^O \cup \hat{\Theta}^I \cup \Omega$ have preferences $\hat{u}^\theta(c_1|\mu_1) > u^\theta(c_0|\mu_1)$, and all student $\theta \in \hat{\Theta}^L$ have preferences $\hat{u}^\theta(c_0|\mu_1) > u^\theta(c_1|\mu_1)$. Given these preferences and the student scores defined above, $\mu_2(c_1) = \mu_0(c_1)$. Therefore, the TIM process cycles, and does not converge.

A similar construction is possible for any N. The preceding logic can be modified such that students $\theta \in \Omega$ are "uniformly at random" likely to most prefer any program $c \in C \setminus \{c_0\}$ for all assignments, and that no student $\theta \notin \Omega$ finds any program $c_j \neq c_1$ preferable to c_0 for any assignment α.

□

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 $\ell = 1,2,3,....$ It begins with all students facing zero peer costs at 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 $N+1$ dimensional 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, $s_1^c := s^c(\lambda(\mu_1)) = s^c(\lambda(\nu_1))$ 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, $s_\ell^c := s^c(\lambda(\mu_\ell)) = s^c(\lambda(\nu_\ell))$ 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.7. *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 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 $\eta(\mu_1(c)) = \eta(\mu_2(c)) < k^c$). Therefore, $s_1^c = s_2^c$.

Induction step: Assume $s_{\ell-1}^c = s_\ell^c$ for all $c \in C_t \setminus D_t^{\ell-1}$. Show $s_\ell^c = s_{\ell+1}^c$ for all $c \in C_t \setminus D_t^\ell$.

No θ with $r^\theta \geq s_\ell^c$ faces peer costs from any program $c \in C_t^\ell$ in matching μ_ℓ. Moreover, by the induction hypothesis, every θ with $r^\theta \geq s_{\ell-1}^c$ faces the same peer costs from any program

$c \in C_t \setminus D_t^{\ell-1}$. Therefore, each θ with $r^\theta \geq s_\ell^c$ 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 $c \in C_t \setminus D_t^\ell$ with scores higher than s_ℓ^c, $\eta(\{\theta \in \mu_\ell(c) | r^\theta > s_\ell^c\}) = \eta(\{\theta \in \mu_{\ell+1}(c) | r^\theta > s_{\ell+1}^c\}) = k^c$ (or 0 if $\eta(\mu_\ell(c)) = \eta(\mu_{\ell+1}(c)) < k^c$). Therefore, $s_\ell^c = s_{\ell+1}^c$.

□

We return to the proof of the proposition.

Proof of stability of μ_t^{SD}

If the terminating step of the algorithm is ℓ', then by construction $\mu_t^{SD} = \mu_{\ell'} = \nu_{\ell'}$ since $D^{\ell'}$ is empty. Therefore, $\mu_t^{SD} = A_t(P_t\lambda(\mu_{\ell'-1}), \lambda(\mu_{\ell'-1}))$, and so $(P_t\lambda(\mu_{\ell'-1}), \lambda(\mu_{\ell'-1}))$ is market clearing. Moreover, because $D^{\ell'}$ is empty it is the case that had we run the algorithm for one more step, we would have had $\mu_{\ell'} = \mu_{\ell'+1}$ 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 $(P_t\lambda(\mu_{\ell'-1}), \lambda(\mu_{\ell'-1}))$ also satisfies rational expectations. By Lemma 2, $\mu_t^{SD} = A_t(P_t\lambda(\mu_{\ell'-1}), \lambda(\mu_{\ell'-1}))$ is stable.

Proof of uniqueness

To show that μ_t^{SD} is the unique stable matching, it suffices to show that $s_{SD} = s(\lambda(\mu_t^{SD}))$ 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, $K = \{c | s_{SD}^c \neq s_*^c\}$. 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 $s^{max} = \max\{s_{SD}^{c_1}, s_*^{c_1}, s_{SD}^{c_2}, s_*^{c_2}, ..., s_{SD}^{c_{|K|}}, s_*^{c_{|K|}}\}$, and let $c^{max} \in \{c | s_{SD}^c = s^{max} \text{ or } s_*^c = s^{max}\}$. 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 $f^\theta(r^\theta, s_{SD}^c) = f^\theta(r^\theta, s_*^c)$ for all $\theta \in I^{max}$ and all $c \in C_t$: each such θ faces the same peer cost from programs c_j 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 $r^\theta > s_{SD}^{c_i}$ 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, and as a result, there cannot exist distinct stable matchings.

Proof of Bullets 1.-3.

1. It suffices to show that there is no step ℓ in the above algorithm such that $c' \in C_t \cap B_2$ 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.7, $s_{SD}^{c'} > s_{SD}^{c}$. 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 4

Proof. We show that in the TIM process, the summary statistics s_t^c of all programs $c \in B_1$ reach s_*^c in finite time. For all t such that $s_t^c = s_*^c$ for all $c \in B_1$, the second part of the claim holds. The following roadmap outlines our proof approach.

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

There is no entry and exit in such markets. The proof is by induction on the index of programs, ordered by their stable statistics in the unique stable matching (Proposition 2).

Suppose $s_*^{c_1} \geq s_*^c$ for all $c \in C$. The **Base Case** finds that there will be a period T 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. Lemma A.8 establishes that if $s_t^{c_1}$ ever falls weakly below $s_*^{c_1}$, then $s_{t'}^{c_1} = s_*^{c_1}$ for all $t' > t$. Lemma A.9 establishes that the maximum summary statistic among all programs cannot always lie above $s_*^{c_1}$. These two claims complete 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**, i.e.** $B_1 \subsetneq C \setminus \{c_0\}$

We extend our results to NSW markets with entry and exit. We argue that the summary statistics of programs in B_1 converge regardless of the entry and exit of programs in B_2, as these programs are always less preferred to programs in B_1 for high-scoring students.

We now begin the proof of the first scenario.

Proof of Scenario 1:

Let $\mu_* = A(p_*, s_*)$ be the unique stable matching, and suppose WLOG that $s_*^{c_1} \geq s_*^{c_2} \geq \ldots \geq s_*^{c_N}$. Note that as there is no exit or entry in the current scenario (as all programs are in B_1), μ_* is the unique stable matching for all t. Let $C' := \{c \in C | s_*^c = 0\}$ be the (possibly empty) subset of programs with a summary statistic of zero in the unique stable matching. We consider the generic case in which at most one program $c \in C'$ has $\eta(\mu_*(c)) = k^c$ 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 $T_1 > 0$ such that for all $t' > T_1$, $s_{t'}^{c_1} = s_*^{c_1}$ and $s_*^{c_1} > s_{t'}^{c_i}$ for all $c_i \neq c_1$.

Proof. The following two lemmas complete the claim.

Lemma A.8. *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, Assumption **AA2** ensures that all students θ with $r^\theta \geq s_*^{c_1}$ satisfy $\succeq^{\theta|s_t} = \succeq^{\theta|s_*}$. Therefore, it must be that (almost) all students θ with $r^\theta \geq s_*^{c_1}$ receive $\mu_{t+1}(\theta) = \mu_*(\theta)$.

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 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 $c_j \neq C_1$. □

Lemma A.8 completes the proof of the Base Case if $s_t^{c_i} \leq s_*^{c_1}$ for all c_i and some $t > 0$. We now argue that must eventually come to pass.

Lemma A.9. *There exists $t > 0$ such that $s_t^{c_i} \leq s_*^{c_1}$ for all c_i.*

Proof. 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 1: 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 the set of students 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 Lemma A.8, all such θ face no peer costs, and therefore $\mu_{t+1}(\theta) = \mu_*(\theta)$ for almost all students θ with $r^\theta \geq s_t^m$. Because $s_{t+1}^m > s_*^{c_1}$ by assumption, no program c_i enrolls enough 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 $s_{t+1}^{c_i} = 0 < s_t^m$ 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. This completes the argument that $s_{t+1}^m < s_t^m$. □

As s_t^m, $t \geq 1$ is a strictly decreasing sequence and $s_t^m \in (s_*^{c_1}, 1]$ for all $t \geq 1$, 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 $k^{c_i} = \eta(\{\theta \in \mu_*(c_i) | r^\theta \geq s_t^m\}) + \eta(\{\theta | r^\theta \in [M_{c_i}^{s_t}, s_t^m)\})$ (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-scoring 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 attend program 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. 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 $s_{t+1}^{c_i}$. 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), $s_{t+1}^{c_i} \leq M_{c_i}^{s_t} < s_t^m - \delta < S$, which contradicts that s_t^m is a decreasing sequence that converges to S.

The remaining possibility is that $S = s_*^{c_1}$, and we procede with this assumption seeking a contradiction. We know that $S > s_*^{c_j}$ for all $j \neq 1$, and so 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 $s_{t'}^{c_1} > S = s_*^{c_1}$. We claim that $s_{t'+1}^{c_1} \leq S$. To see this, note that students θ with scores $r^\theta \in [s_{t'}^{c_1}, 1]$ receive $\mu_{t'+1}(\theta) = c_1$ if and only if $\mu_*(\theta) = c_1$ but some students with $r^\theta \in [S, s_{t'}^{c_1})$ who receive $\mu_*(\theta) = c_1$ may not receive $\mu_{t'+1}(\theta) = c_1$ because they face peer costs. Therefore, it must be that $s_{t'+1}^{c_1} \leq S$. Contradiction. □

□

Induction Step: Suppose $c_j \notin C'$ and that there exists some time $T_{j-1} > 0$ such that for all $t' \geq T_{j-1}$ all programs c_i, $i < j$ have $s_{t'}^{c_i} = s_*^{c_i}$ and all programs c_i, $i > j$ have $s_*^{c_i} \leq s_*^{c_{j-1}}$. Then there exists T_j such that for all $t'' > T_j$, $s_{t''}^{c_j} = s_*^{c_j}$ 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.

1. If there is some time $T_j - 1 > 0$ such that $s_{T_j-1}^{c_i} \leq s_*^{c_j}$ for every c_i with $i \geq j$ and $s_{T_j-1}^{c_i} = s_*^{c_i}$ for all c_i with $i < j$, then for all $t'' > T_j - 1$, $s_{t''}^{c_j} = s_*^{c_j}$ and $s_*^{c_j} > s_{t''}^{c_i}$ for all $i < j$.
2. Let $s_t^{m,j} = \max_{c_i, i \geq j} s_t^{c_i}$. Then if $s_t^{m,j} > s_*^{c_j}$ for all $t \geq T_{j-1}$, $s_{T_{j-1}}^{m,j} > s_{T_{j-1}+1}^{m,j} > \ldots$.
3. There exists some $T_j - 1$ such that $s_{T_j-1}^{m,j} \leq s_*^{c_j}$.

The argument for the first claim is completely analogous to that in Lemma A.8, with the change in notation of s_t^m to $s_t^{m,j} = \max_{c_i, i \geq j} s_t^{c_i}$. The arguments for the second and third claims hold while noting that the analogous arguments in Lemma A.9 hold while fixing student preferences given $s_*^{c_i}$ for all time $t > T_j - 1$ for all programs c_i, $i < j$. □

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 $T - 1 > 0$ such that $s_{T-1}^{c_j} = 0$ for every $c_j \in C$ then $s_T^{c_j} = 0$.
2. Let $s_t^{m,0} = \max_{c_j \in C'} s_t^{c_j}$ for $t \geq T$. Then $s_T^{m,0} > s_{T+1}^{m,0} > \ldots$.
3. There exists some $T - 1$ such that $s_{T-1}^{m,0} = 0$.

To see that the first claim holds, note that if $c_i \in C'$ then $c_j \in C'$ for any $j > i$. Therefore all programs with sufficiently high indices have stable summary statistics equal to 0, and then, by an argument analogous to Lemma A.8, they converge to their stable value immediately. The argument for the second claim is analogous to Claim 1 of Lemma A.9.

The third claim is shown by a similar argument as in Lemma A.9. Given that $s_t^{m,0}$ is bounded and decreasing, it must converge. The argument in Lemma A.9 establishes that $s_t^{m,0}$ 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 guarantees that there is at most one program $c_J \in C'$ for which $\eta(\mu(c_J)) = k^{c_J}$, and for all other programs $c_j \in C'$, $\eta(\mu(c_j)) < k^{c_j}$. The remainder of the argument is nearly identical to the "δ" argument in Lemma A.9. □

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, i.e. $B_1 \subsetneq C \setminus \{c_0\}$

Proof. We claim that entry and exit of programs in B_2 do not affect the convergence property of the TIM process for top-block programs by arguing that the proofs of Lemmas A.8 and A.9 are largely unaltered for programs in B_1. The remainder of the argument (i.e. induction step) follows as in **Scenario 1**.

In any sequence of NSW markets $\{E_t\}_{t \geq 1}$, we know from Proposition 2 that the associated sequence of stable matchings $\{\mu_t^*\}_{t \geq 1}$ is such that for all t and all $c \in B_1$, $s^c(\lambda(\mu_t^*))$ is a constant value, which we denote by s_*^c. Furthermore, for all $c' \in B_2$ and all t, $s_*^c \geq s^{c'}(\lambda(\mu_t^*))$.

Consider the proof of Lemma A.8. By **AA2**, if $s_t^{c_i} < s_*^{c_1}$ for all $c_i \in B_1$ then (almost) all θ such that $r^\theta > s_t^m$ will receive $\mu_{t+1}(\theta) = \mu_*(\theta)$ (where we do not index $\mu_*(\theta)$ with t following the result of Proposition 2). By definition, this means that there will be exactly k^{c_1} students with scores $r^\theta > s_*^{c_1}$. Therefore, $s_{t+1}^{c_1} = s_*^{c_1}$ and $s_{t+1}^{c_i} < s_*^{c_1}$ for all $c_i \in B_1 \setminus \{c_1\}$.

Consider the proof of Lemma A.9. The proof for Claim 1 is unaltered, where s_t^m is now defined as $\max_{c_i \in B_1} s_t^{c_i}$. The remaining argument is also unchanged; a program c entering at any period t can only have the effect of lowering $s_t^{c_i}$ for $c_i \in B_1$ compared to the counterfactual in which c did not enter at time t. The exit of a program c at time t does not have any effect on the matching $\mu(\theta)_t$ for (almost) all θ with $r^\theta \geq s_t^m$ by Assumption **AA2**. $M_{c_i}^{s_t}$ does not depend on the set of programs present in the market, and is an upper bound on $s_{t+1}^{c_i}$ for all $c_i \in B_1$. Therefore, the remainder of the argument in the proof of the lemma is also unchanged.

□

Remark 4

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 **AA3**.

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 Let $\alpha = A(p,\lambda)$ and $\alpha' = A(p',\lambda')$.

First consider the case in which $s(\lambda(\alpha)) = p^c$ and $s(\lambda(\alpha')) = p'^c$ for some $c \in C$. Then the proof follows from Lemma A.3.

Second, consider the case in which (without loss of generality) $s(\lambda(\alpha)) > p^c$ and $s^c(\lambda(\alpha)) - s^c(\lambda(\alpha')) = \gamma > 0$ for some program $c \in C$. Because $s(\lambda(\alpha)) > p^c$, by Assumption **AA3** we know that $\eta(\{\theta \in \alpha(c) | r^\theta > s^c(\lambda(\alpha))\}) = k^c$, and therefore $\eta(\alpha(c)) > k^c$. For any δ, note that $||\lambda(\alpha) - \lambda(\alpha')||_\infty < \delta$ implies that $|\eta(\alpha(c)) - \eta(\alpha'(c))| = |\lambda^{c,1}(\alpha) - \lambda^{c,1}(\alpha')| < \delta$. Therefore, for sufficiently small δ, it must be that $\eta(\{\theta \in \alpha'(c) | r^\theta > s^c(\lambda(\alpha'))\}) = k^c$, i.e. $s(\lambda(\alpha')) > p'^c$. Then the following holds for sufficiently small δ

$$\begin{aligned}
\delta &> |\lambda^{c,1}(\alpha) - \lambda^{c,1}(\alpha')| \\
&= |\lambda^{c,1}(\alpha) - k^c - \lambda^{c,1}(\alpha') + k^c| \\
&= |\lambda^{c,1}(\alpha) - [\eta(\alpha(c)) - \lambda^{c,s^c(\lambda(\alpha))}(\alpha)] - (\lambda^{c,1}(\alpha') - [\eta(\alpha'(c) - \lambda^{c,s^c(\lambda(\alpha'))}(\alpha')])| \\
&= |\lambda^{c,s^c(\lambda(\alpha))}(\alpha) - \lambda^{c,s^c(\lambda(\alpha'))}(\alpha')|
\end{aligned} \tag{A.8}$$

where the second equality follows from the earlier argument that $\eta(\{\theta \in \alpha(c) | r^\theta > s^c(\lambda(\alpha))\}) = k^c$ and $\eta(\{\theta \in \alpha'(c) | r^\theta > s^c(\lambda(\alpha'))\}) = k^c$, and the final equality follows from $\eta(\alpha(c)) = \lambda^{c,1}(\alpha)$ and $\eta(\alpha'(c)) = \lambda^{c,1}(\alpha')$.

By the ongoing assumption that $||\lambda(\alpha) - \lambda(\alpha')||_\infty < \delta$, it must be that

$$|\lambda^{c,s^c(\lambda(\alpha))}(\alpha) - \lambda^{c,s^c(\lambda(\alpha))}(\alpha')| < \delta. \tag{A.9}$$

Recall that an earlier step in the proof of this result has established that **A1** holds. Therefore, we complete the proof by arguing that for any sufficiently small δ, $\gamma < \frac{2\delta}{\omega}$ for some constant $\omega > 0$. **A2** implies that $\lambda^{c,s^c(\lambda(\alpha))}(\alpha') > \lambda^{c,s^c(\lambda(\alpha'))}(\alpha') + \gamma\omega$. But then jointly satisfying Inequalities A.8 and A.9 require that $\gamma\omega < 2\delta$. Rearranging yields $\gamma < \frac{2\delta}{\omega}$ as desired.

□

Local Convergence of TIM

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

Definition 7. *A stable matching $\mu_* = (p_*, \lambda_*)$ is* locally convergent in the TIM process *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, because we restrict ourselves to initial 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 5. *A stable matching μ_* is not necessarily locally convergent in the TIM process, even if it is the unique stable matching in market E.*

Proof. By Theorem 2, it suffices to find a market such that $\lambda_0,\lambda_1,...$ does not converge for any $\lambda_0 \neq \lambda_* = \lambda(\mu_*)$, where μ_* is the unique stable matching. The following example is such a market.

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

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

Each θ receives zero utility from remaining unmatched. $\gamma < 1$ measure 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" over $[0,1]$. The remaining $1-\gamma$ measure of students have strong peer preferences and receive utility $v^\theta - f(s(\lambda),r^\theta)$ from matching with the program, where

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

for some $K > 0$ and each v^θ is 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. Any θ is better off enrolling at the program if and only if $v^\theta - f(s(\lambda(\alpha)),r^\theta) \geq 0$, where we break ties in favor of the student attending the program.

We claim that such that $\mu_*(\theta) = c$ for all $\theta \in \Theta$ is the unique stable matching. μ_* is a matching since $q^c \geq 1$. Then $\lambda_* = \lambda(\mu_*)$ has the property that $\lambda_*^{c,(y,y)} = y$ for all $y \in [0,1]$. 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 process 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 students 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 $s_{t+1}, s_{t+3}, s_{t+5}$,... is non-increasing where each element is strictly smaller than $\frac{1}{2}$. Therefore, the TIM process does not converge.

□

Proposition 3

Proof.

1. If the TIM process converges to $\mu_* = A(p_*, \lambda_*)$ in market E given μ_0, then for any stopping rule $\delta > 0$ the TFM mechanism must terminate in market E given μ_0, 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 process, 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 process 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 λ_1, λ_2,... must be Cauchy because it is convergent. Let $\delta \in (0, \min_{\tau \leq \tau(\gamma)} \Delta_\tau)$. 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 $\mu_{\mu_0,\delta}(\theta) = A(P\lambda_{\tau(\delta)-1}, \lambda_{\tau(\delta)-1})$, and recall that $\mu_* = A(P\lambda_*, \lambda*)$. By Lemma A.3 this implies that for any $\epsilon > 0$ there exists $\delta^* < \min_{\tau \leq \tau(\gamma)} \Delta_\tau$ such that for any stopping rule $\delta < \delta^*$, $\eta(\{\theta | \mu_{\mu_0,\delta}(\theta) = \mu_*(\theta)\}) > 1 - \epsilon$.

2. Fix $\epsilon > 0$, $\delta > 0$, and a stable matching μ_*. The proof of point 1 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 stopping criterion δ if $||\lambda_* - \lambda_0||_\infty < \min\{\gamma_1, \gamma_2\}$.

   Example 4 constructs a market E in which the TIM process will not converge for any $\lambda_0 \neq \lambda_*$, thus completing the claim.

3. Suppose the TFM mechanism terminates in period $\tau(\delta) > 0$. Because the final matching is not constructed at any step $\tau < \tau(\delta)$ in which λ_τ is being updated, and because each λ_τ is unaffected by the submitted preferences of any zero measure set of student, no student affects the final matching by misreporting preferences in any step $\tau < \tau(\delta)$. 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 μ_0 and $\epsilon > 0$. Termination of the TFM mechanism implies that $||\lambda_{\tau(\delta)} - \lambda_{\tau(\delta)-1}||_\infty < \delta$. Assuming (almost) all students $\theta' \in \Theta$ report preferences $\succeq^{\theta'|\mu_{\tau(\delta)-1}}$, we have that any $\theta \in \Theta$ can profitably misreport her preferences only if $\succeq^{\theta|\mu_{\tau(\delta)}} \neq \succ^{\theta|\mu_{\tau(\delta)-1}}$. By Assumption **A4**, there exists δ^* such that $\eta(\{\theta | \succeq^{\theta|\mu_{\tau(\delta)}} \neq \succ^{\theta|\mu_{\tau(\delta)-1}}\}) < \epsilon$ for any stopping rule $\delta < \delta^*$. Therefore, $\eta(\Theta') < \epsilon$ for sufficiently small δ, as desired.

4. Fix $\epsilon > 0$ and (μ_0, δ). Suppose that the TFM mechanism terminates at step $\tau(\delta) = K \cdot T + t$. Note that the stopping criterion is independent of K, T, t, i.e. $\tau(\delta)$ is a constant. There exists T_1 such that for any $T > T_1$, $K = 0$ and $\tau(\delta) = t$. Moreover, for any $\gamma > 0$ there exists $T_2 > T_1$ such that $\frac{t}{T} = \frac{\tau(\delta)}{T} < \gamma$ for any $T > T_2$. $K = 0$ implies that no student reports her preferences more than twice, and $t = \tau(\delta)$ implies that the share of submarkets who report preferences twice is $\frac{\tau(\delta)}{T}$. Recall our assumption that $\eta(\Theta_\ell) \to 0$ for all $\ell \in 1, ..., T$ as $T \to \infty$. Therefore, there exists T^* such that the measure of students asked to report preferences twice is given by

$$\sum_{\ell=1}^{\tau(\delta)} \eta(\Theta_\ell) < \epsilon$$

   for any $T > T^*$.

□

C Additional Theoretical Results on the TIM Process

Let α and α' be any two assignments. Let $\eta \in \hat{H}(N)$ and suppose that $f \mapsto \eta$. We say that η *exhibits weakly increasing peer preferences in size* if $f^{\theta,c}(\alpha(c)) \geq f^{\theta,c}(\alpha'(c))$ for almost all θ and any $c \in C \setminus \{c_0\}$

such that $\eta(\alpha(c)) \geq \eta(\alpha'(c))$. We say that $\eta \in \hat{H}(N)$ *exhibits strictly increasing peer preferences in size* if $f^{\theta,c}(\alpha(c)) > f^{\theta,c}(\alpha'(c))$ for almost all θ and any $c \in C \setminus \{c_0\}$ such that $\eta(\alpha(c)) > \eta(\alpha'(c))$.

Remark 6. *If $\eta \in \hat{H}(N)$ exhibits weakly increasing peer preferences in size then it admits no negative externality groups.*

Proof. Suppose $\eta \in \hat{H}(N)$ exhibits weakly increasing peer preferences in size, and suppose for contradiction that η admits a negative externality group. Then there exists an assignment α, a program $c \in C \setminus \{c_0\}$, and positive measure sets of students $\Theta^I \subset \alpha(c)$ and $\Theta^O \subset \Theta \setminus \alpha(c)$ with $\eta(\Theta^I) \geq \eta(\Theta^O)$ such that $f^{\theta,c}(\lambda^c(\Theta^O \cup \alpha(c) \setminus \Theta^I)) > f^{\theta,c}(\lambda^c(\alpha(c)))$ for all $\theta \in \Theta^I$. Then $f^{\theta,c}(\cdot)$ is not weakly increasing in size for all $\theta \in \Theta^I$ since $\eta(\Theta^O \cup \alpha(c) \setminus \Theta^I) \leq \eta(\alpha(c))$. Contradiction.

□

Proposition 4.

1. *For any $N < 3$ let $\eta \in \hat{H}(N)$. If η exhibits weakly increasing peer preferences in size, then the TIM process converges for any starting condition μ_0 in any market $E = [\eta, q, N]$.*

2. *For any $N \geq 3$ let $\eta \in \hat{H}(N)$ and suppose that $f \mapsto \eta$. If η exhibits strictly increasing peer preferences in size then there exists a market $E = [\hat{\eta}, q, N]$ such that $f \mapsto \hat{\eta}$ and a starting condition μ_0 for which the TIM process does not converge in market E.*

Proof. In what follows, we write $s_t^c := \eta(\alpha(\mu_t(c))$ where μ_t, $t \geq 0$ is the time t matching in the TIM process. Note that by Theorem 2 it suffices to investigate the convergence (or lack of convergence) of s_t^c in t for all $c \in C \setminus \{c_0\}$.

1. First suppose $N = 1$. It is either the case that $s_1^{c_1} \geq s_0^{c_1}$ or $s_1^{c_1} < s_0^{c_1}$. If $s_1^{c_1} \geq s_0^{c_1}$, then for almost every $\theta \in \mu_1(c_1)$, it is also the case that $\theta \in \mu_2(c_1)$ since almost all students find the program weakly more attractive as its size increases. By repeated application of this argument, $s_t^{c_1}$ is a weakly increasing sequence. Similarly, if $s_1^{c_1} < s_0^{c_1}$, then $s_t^{c_1}$ is a weakly decreasing sequence, as the program becomes less attractive to almost all students as it shrinks in size. In either case, $s_t^{c_1}$ is a monotonic sequence on a bounded set; for any t, $s_t^{c_1} \in [\omega, q^{c_1}]$ where ω is a "full support" constant assumed in **A2**. Therefore, $s_t^{c_1}$ converges to some value $s \in [\omega, q^{c_1}]$.

   Now suppose $N = 2$. If $s_t^{c_1}$ and $s_t^{c_2}$ are both weakly increasing or both weakly decreasing in t, then we similarly have convergence. Therefore, suppose there is some period $T_1 \geq 0$ such that WLOG $s_{T_1+1}^{c_1} \geq s_{T_1}^{c_1}$ and $s_{T_1+1}^{c_2} \leq s_{T_1}^{c_2}$, with at least one inequality strict.

   Claim 1: We claim that $s_t^{c_1} \leq s_{t+1}^{c_1}$ and $s_t^{c_2} \geq s_{t+1}^{c_2}$ for all periods $t \in \{T_1, ..., T_2 - 1\}$, where T_2 is the first period t such that $s_t^{c_1} = q^{c_1}$ if such a time period exists, and $T_2 = \infty$ otherwise. The reason for this is the same as in the $N = 1$ case: c_1 is becoming increasingly more attractive to almost all students, and c_2 is becoming increasingly less attractive to almost all students. At each period,

a weakly larger measure of workers will join c_1 and a weakly fewer measure of workers will join c_2, and capacity constraints do not bind prior to period T_2.

If $T_2 = \infty$, then we have established convergence, as $s_t^{c_2} \geq s_{t+1}^{c_2}$ and $s_t^{c_1} \leq s_{t+1}^{c_1}$ for all $t \geq T_1$. Therefore, we proceed with the assumption that $T_2 < \infty$.

Claim 2: If $T_2 < \infty$ and $s_{T_2+1}^{c_2} \leq s_{T_2}^{c_2}$, then $s_t^{c_2} \leq s_{t+1}^{c_2}$ for all $t > T_2$. Recall that $s_{T_2}^{c_2} \leq s_{T_2-1}^{c_2}$ and $q^{c_1} = s_{T_2}^{c_1} \geq s_{T_2-1}^{c_1}$. There are two cases of interest. First, if $s_{T_2+1}^{c_2} = s_{T_2}^{c_2} = q^{c_2}$, then by Theorem 2, we have converged to a stable matching. Second, if $s_{T_2+1}^{c_2} \leq s_{T_2}^{c_2} \leq q^{c_2}$ with at least one strict inequality, then almost every student $\theta \in \mu_{T_2}(c_0)$ finds c_2 unacceptable when $s^{c_2} \leq s_{T_2+1}^{c_2}$. Moreover, any subset of students $\Theta' \subset \mu_{T_2+1}(c_2)$ can displace at the largest an equal measure of students $\Theta'' \subset \mu_{T_2+1}(c_1)$. Therefore, $s_{T_2+2}^{c_2} \leq s_{T_2+1}^{c_2}$, and repeated application of this logic verifies the claim.

By Claim 2, the remaining case is therefore that $T_2 < \infty$ and $s_{T_2+1}^{c_2} > s_{T_2}^{c_2}$.

Claim 3: Suppose there exists some time period $t > T_2$ such that $s_t^{c_1} < q^{c_1}$, and let T_3 be the first such period. Then $s_{T_3-1}^{c_2} > s_{T_1}^{c_2}$. To see this, suppose for contradiction that $s_{T_3-1}^{c_2} \leq s_{T_1}^{c_2}$. Because $s_{T_3-1}^{c_1} = q^{c_1}$ by construction, c_1 is weakly more attractive for almost every student at time T_3 than it was at time T_1+1. Because $s_{T_3-1}^{c_2} \leq s_{T_1}^{c_2}$, c_2 is weakly less attractive for almost every student at time T_3 than it was at time T_1+1. Therefore, it must be the case that $s_{T_3}^{c_1} > s_{T_3-1}^{c_1} = q^{c_1}$. Contradiction with T_3 being the first time period $t > T_2$ such that $s_t^{c_1} < q^{c_1}$.

Therefore, we continue by supposing there exists some first period $T_3 > T_2$ such that $s_{T_3}^{c_1} < q^{c_1}$ and $s_{T_3-1}^{c_2} > s_{T_1}^{c_2}$. By Claim 1, it must be the case that $s_t^{c_2} \leq s_{t+1}^{c_2}$ and $s_t^{c_1} \geq s_{t+1}^{c_1}$ for all periods $t \in \{T_3-1,...,T_4-1\}$, where T_4 is the first period t such that $s_t^{c_2} = q^{c_2}$ if such a time period exists, and $T_4 = \infty$ otherwise. Moreover, it must be the case that $s_{t+1}^{c_2} \geq s_t^{c_2}$ for all $t \geq T_3 - 1$; by Claim 3, it cannot be the case that $s_t^{c_2} < q^{c_2}$ for any $t > T_4$ because this would require $s_t^{c_1} > s_{T_3-1}^{c_1} = q^{c_1}$ which is impossible as c_1 cannot exceed its capacity in any matching. Therefore, $s_t^{c_1}$ is a nondecreasing sequence for all $t \geq T_4$, implying that it must converge.

It remains to argue then that $s_t^{c_1}$ must converge. If $T_4 = \infty$, then $s_{t+1}^{c_1} \geq s_t^{c_1}$ for all $t \geq T_3$ and must therefore converge. If $T_4 < \infty$ then by Claim 2, if $s_{T_4+1}^{c_1} \leq s_{T_4}^{c_1}$, then $s_{t+1}^{c_1} \geq s_t^{c_1}$ for all $t \geq T_4$ and must therefore converge. If $T_4 < \infty$ and $s_{T_4+1}^{c_1} > s_{T_4}^{c_1}$, then c_1 is weakly more attractive to almost all students at $T_4 + 2$ than at $T_4 + 1$, and because $s_t^{c_2} = q^{c_2}$ for any $t > T_4$, it must be the case that $s_{T_4+2}^{c_1} \geq s_{T_4+1}^{c_1}$, and repeated application of this logic implies that $s_{t+1}^{c_1} \geq s_t^{c_1}$ for all $t > T_4$ and therefore must converge.

2. Let $N = 3$, i.e. $C = \{c_0, c_1, c_2, c_3\}$. Each student's payoff from attending a program is weakly increasing in the measure of other students attending that program. Formally, the utility attained by student θ from being matched to program c at assignment α is $u^\theta(c|\alpha) = v^{\theta,c} + f^{\theta,c}(\eta(\alpha(c))$ where $f(\cdot)$ is weakly increasing.

We represent the relevant preference and score restrictions in the following two tables. For notational compactness, we partition the set of students into twelve disjoint groups $\{G_1,G_2,...,G_{12}\}$. We write "students in group G_ℓ prefer program c_m to program c_p to all other programs if $\eta(\alpha(c_m)) > x$" for some value $x \in [0,1]$ as "$c_m \succ_{G_l} c_p \succ_{G_l} c_R$ if $\eta(\alpha(c_m)) > x$" where we omit the subscript in the "$\succ$" term when the group identity is clear. To characterize relative scores, we use the notation $r^{G_\ell,c} > r^{G_j,c}$ to represent that for all students $\theta \in G_\ell$ and all $\theta' \in G_j$, $r^{\theta,c} > r^{\theta',c}$. In words, all students in G_ℓ have higher scores than any student in G_j for program c.

We also represent the vector of measures of students enrolled at each program by $s = (s^{c_1},s^{c_2},s^{c_3}) := (\eta(\alpha(c_1),\eta(\alpha(c_2),\eta(\alpha(c_3))$, as this is a sufficient statistic for student preferences.

Group	**Measure**	**Preference Restrictions**
G_1	0.015	$c_3 \succ c_0 \succ c_R \ \forall \alpha$
G_2	0.04	$c_3 \succ c_2 \succ c_R$ if $\eta(\alpha(c_3)) > 0.193$, $c_2 \succ c_3 \succ c_R$ if $\eta(\alpha(c_3)) < 0.192$
G_3	0.02	$c_2 \succ c_R$ if $\eta(\alpha(c_2)) > 0.226$, $c_3 \succ c_R$ if $\eta(\alpha(c_2)) < 0.225$ and $\eta(\alpha(c_1)) < 0.196$, $c_1 \succ c_R$ if $\eta(\alpha(c_1)) > 0.197$ and $\eta(\alpha(c_2)) < 0.225$
G_4	0.01	$c_1 \succ c_2 \succ c_R$ if $\eta(\alpha(c_1)) > 0.216$ and $\eta(\alpha(c_2)) < 0.201$, $c_2 \succ c_1 \succ c_R$ if $\eta(\alpha(c_1)) < 0.215$ or $\eta(\alpha(c_2)) > 0.202$
G_5	0.045	$c_1 \succ c_3 \succ c_R$ if $\eta(\alpha(c_1)) > 0.214$, $c_3 \succ c_1 \succ c_R$ if $\eta(\alpha(c_1)) < 0.213$
G_6	0.01	$c_1 \succ c_3 \succ c_R$ if $\eta(\alpha(c_1)) > 0.205$, $c_3 \succ c_1 \succ c_R$ if $\eta(\alpha(c_1)) < 0.204$
G_7	0.38	$c_0 \succ c_R \ \forall \alpha$
G_8	0.16	$c_1 \succ c_R \ \forall \alpha$
G_9	0.19	$c_2 \succ c_R \ \forall \alpha$
G_{10}	0.13	$c_3 \succ c_R \ \forall \alpha$
G_{11}	0.005	$c_1 \succ c_3 \succ c_R$ if $\eta(\alpha(c_1)) > 0.214$, $c_3 \succ c_1 \succ c_R$ if $\eta(\alpha(c_1)) < 0.213$
G_{12}	0.005	$c_3 \succ c_0 \succ c_R \ \forall \alpha$

Program	Capacity	Score Restrictions
c_0	1	None
c_1	0.5	None
c_2	0.5	None
c_3	0.2	$r^{G_{10},c_3}, r^{G_2,c_3}, r^{G_3,c_3}, r^{G_6,c_3} > r^{G_1,c_3}, r^{G_5,c_3}, r^{G_{11},c_3}, r^{G_{12},c_3}$ $r^{G_1,c_3} > r^{G_5,c_3}$ $r^{G_{11},c_3} > r^{G_{12},c_3}$

A4 is not contradicted by the above restrictions as we define each group's preferences contingent on program size over disconnected halfspaces in $[0,1]^4$. **A2** is not contradicted by the above restrictions for the following reasons. First, note that there exist positive measure sets of students that most prefer each program regardless of α. For programs c_1 and c_2 we place no constraints on student scores, therefore, and we can therefore satisfy **A2** by "uniformly distributing" the scores of all students. We can also "uniformly distribute" the scores at c_3 of students in groups G_4, G_7, G_8, G_9 over $[0,1]$. Similarly, note that students in G_1, G_{10}, and G_{12}. Because $\eta(G_2 \cup G_3 \cup G_{10} \cup G_6) = 0.2$, $\eta(G_1 \cup G_5 \cup G_{11} \cup G_{12}) = 0.07$, and $\eta(G_1 \cup G_{11}) = 0.02$ so we can "uniformly distribute" scores of G_2, G_3, G_{10}, G_6 at c_3 over $[1-\frac{0.2}{0.27},1]$, "uniformly distribute" scores of G_1, G_{11} at c_3 over $[1-\frac{0.22}{0.27},1-\frac{0.2}{0.27})$, and "uniformly distribute" scores of G_5, G_{12} at c_3 over $[0,1-\frac{0.22}{0.27})$. Therefore, "uniformly distributed" over each of these intervals is a positive measure set of students who most prefers c_3 for any assignment.

At period some $T > 0$ the matching μ_T generated in the TIM process is such that $\eta(\mu_T(c_1)) = \eta(\mu_T(c_2)) = \eta(\mu_T(c_3)) = 0.2$, so that $s_T = (0.2,0.2,0.2)$. The following table shows the matchings created in the TIM process for periods $T,...,T+6$. For the above set of preferences and scores, the sequence of matchings cycle, and we have $\mu_t = \mu_{t+7}$ for any $t \geq T$, even though these two matchings are different from μ_{t+1}, μ_{t+2}, μ_{t+3}, μ_{t+4}, μ_{t+5}, and μ_{t+6}. Therefore, the TIM process does not converge.

We will represent the matchings showing which groups are enrolled at each program in the following table. Given that G_7, G_8, G_9 and G_{10} are always enrolled, respectively, at c_0,c_1,c_2,c_3, we exclude them from it. We use a bold font whenever the group moves to a certain program in that period.

c	μ_T	μ_{T+1}	μ_{T+2}	μ_{T+3}	μ_{T+4}	μ_{T+5}	μ_{T+6}
c_0	G_1,G_{12}	G_{12}					$\mathbf{G_{12}}$
c_1	G_5,G_{11}	$\mathbf{G_3}$,G_5	G_3,$\mathbf{G_4}$,$\mathbf{G_5}$,$\mathbf{G_6}$,$\mathbf{G_{11}}$	G_3,G_4,G_5,G_6,G_{11}	G_5,G_6,G_{11}	G_6	$\mathbf{G_5}$
c_2	G_4	G_4		$\mathbf{G_2}$	G_2,$\mathbf{G_3}$,$\mathbf{G_4}$	G_2,G_3,G_4	G_3,G_4
c_3	G_2,G_3,G_6	$\mathbf{G_1}$,G_2,$\mathbf{G_{11}}$,G_6	$\mathbf{G_1}$,$\mathbf{G_{12}}$,G_2	G_1,G_{12}	G_1,G_{12}	G_1,G_{12},$\mathbf{G_5}$,$\mathbf{G_{11}}$	G_1,$\mathbf{G_2}$,G_{11},$\mathbf{G_6}$

Finally, note that our construction of preferences can be satisfied for any collection of strictly increasing functions $\{f^{\theta,c}(\cdot)\}_{\theta\in\Theta,c\in C\setminus\{c_0\}}$ by appropriately selecting $\{v^{\theta,c}\}_{\theta\in\Theta,c\in C}$. Also, similarly to bullet 3 of Theorem 3, this example can be straightforwardly extended to $N \geq 4$ programs.

□

D Additional Theoretical Results in New South Wales Markets

We have shown that the summary statistics of top-block programs converge in finite time in any NSW market. However, 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 $t>T$. The "big-fish" preferences present in NSW markets imply that even these students do not receive a negative utility shock to their preferences. We say that a student θ is a member of a *negative utility blocking pair* at time $t \geq 1$ in the TIM process if there exists $c \in C_t$ such that (θ,c) form a blocking pair, and $u^\theta(\mu_t(\theta)|\mu_{t-1}) > u^\theta(\mu_t(\theta)|\mu_t)$. The following summarizes this statement, and is presented without proof.

Remark 7. *In a generic NSW market, the measure of students involved in negative utility blocking pairs goes to zero in the TIM process, that is,*
$\eta(\{\theta \in \bigcup_{c\in B_1} \mu_t(c) | \theta \text{ is a member of negative utility blocking pair}\} \to 0$. *Moreover, if* $s_c^* = 0$ *for at most one program* $c \in B_1$, *then there exists some time* $T<\infty$ *such that*
$\eta(\{\theta \in \bigcup_{c\in B_1} \mu_t(c) | \theta \text{ is a member of negative utility blocking pair}\} = 0$ *for all* $t>T$ *in the TIM process.*

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.

We can also study how the rate of entry and exit affects the amount of instability in the market. 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 μ_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 the 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. For notational simplicity, we assume that $B_1 = C \setminus \{c_0\}$ and show that the TIM process converges from any starting condition μ_0 in no more than $N+2$ periods. Therefore, if entry or exit occurs far less often than once every $N+2$ periods, the TIM process will (in most periods) generate a stable matching.

Remark 8. *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* $t > N+1$.

Proof. Let $\epsilon \in (0,1)$ and let E^ϵ be a NSW market where $1-\epsilon$ measure of students have common intrinsic preferences, that is $\eta\{\theta | v^{\theta,c_1} > v^{\theta,c_2} > ... > v^{\theta,c_N}\} = 1-\epsilon$. Let $\tilde{E}^\epsilon$ be a market that differs from E^ϵ only in

that we permute student intrinsic preference such that $v^{\theta,c_1} > v^{\theta,c_2} > ... > v^{\theta,c_N}$ 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 $\delta_1 > 0$ there exists $\epsilon' > 0$ such that for all $\epsilon < \epsilon'$, $||s_* - \tilde{s}_*||_\infty < \delta_1$.

Step 2: For any $i \neq j$ such that $\tilde{s}_*^{c_i} > 0$, there exists $\delta_2 > 0$ such that $|\tilde{s}_*^{c_i} - \tilde{s}_*^{c_j}| > \delta_2$. Similarly, there exists ϵ' such that for any $\epsilon < \epsilon'$, $|s_*^{c_i} - s_*^{c_j}| > \delta_2$.

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 $\delta_4 > 0$ and μ_0, there exists $\epsilon' > 0$ such that for $\epsilon < \epsilon'$, $||\tilde{s}_t - s_t||_\infty < \delta_4$ for all $t \leq 3N+1$ in the TIM process.

Step 5: For any μ_0 and $\delta_5 > 0$ there exists $\epsilon' > 0$ such that for any $\epsilon < \epsilon'$, the TIM process in E^ϵ yields s_{N+1} such that $||s_{N+1} - s_*||_\infty < \delta_5$. Continuing on in the TIM process, $s_{3N+1} = s_*$.

We now prove each step in the order presented:

Step 1: For any $\delta_1 > 0$ there exists $\epsilon' > 0$ such that for all $\epsilon < \epsilon'$, $||s_* - \tilde{s}_*||_\infty < \delta_1$.

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 at $t = 1$ in markets E^ϵ and $\tilde{E}^\epsilon$ (Lemma A.3). By Remark 4, for sufficiently small γ_1, $|s_1^{c_1} - \tilde{s}_1^{c_1}| < \delta_1$, 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 markets 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 $\delta_2 > 0$ such that $|\tilde{s}_*^{c_i} - \tilde{s}_*^{c_j}| > \delta_2$. Similarly, there exists ϵ' such that for any $\epsilon < \epsilon'$, $|s_*^{c_i} - s_*^{c_j}| > \delta_2$.

Proof. If $\tilde{s}_*^{c_i} > 0$ then $\tilde{s}_*^{c_j} > \tilde{s}_*^{c_i}$ for $j < i$. This follows because $v^{\theta,c_j} > v^{\theta,c_i}$ for almost all θ, and therefore at most a measure zero 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 $\delta_2 \in (0, \min_{i \in N'} s_*^{c_i} - s_*^{c_{i+1}})$ satisfies our requirement.

That there exists ϵ' such that for any $\epsilon < \epsilon'$, $s_*^{c_i} - s_*^{c_{i+1}} > \delta_2$ 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 $j < i \leq N+1$ 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 $\{\theta | r^\theta \geq \tilde{s}_*^{c_i}\}$. Any such θ will have $\tilde{\mu}_{t+1}(\theta) = c_\ell$ for $\ell \leq i$ by Assumption **AA2** and the fact that intrinsic preferences are fully aligned in market $\tilde{E}^\epsilon$. Therefore, $\tilde{s}_t^{c_\ell} \leq \tilde{s}_*^{c_i}$ for all $\ell > i$. By **Scenario 1** from the proof of Theorem 4, 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, $\tilde{s}_1^{c_1} \leq \tilde{s}_*^{c_1}$. Therefore, 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 $j < i \leq N+1$ 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 $i+1 \leq N+1$.

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

□

Step 4: For any $\delta_4 > 0$ and μ_0, there exists $\epsilon' > 0$ such that for $\epsilon < \epsilon'$, $||\tilde{s}_t - s_t||_\infty < \delta_4$ for all $t \leq 3N+1$.

Proof. Fix μ_0 and $\delta_4 > 0$. Define μ_t and $\tilde{\mu}_t$ as the matchings formed at t for markets E^ϵ and $\tilde{E}^\epsilon$, respectively. By Lemma A.3 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, $||s_1 - \tilde{s}_1||_\infty < \delta_4$. By repeated application of Lemma A.3, 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, $||s_t - \tilde{s}_t||_\infty < \delta_4$. To complete the result, let $\epsilon' = \min_{t \leq 3N+1} \epsilon_t$.

□

Step 5: For any μ_0 and $\delta_5 > 0$ there exists $\epsilon' > 0$ such that for any $\epsilon < \epsilon'$, the TIM process in E^ϵ yields s_{N+1} such that $||s_{N+1} - s_*||_\infty < \delta_5$. 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, $s_t^{c_i} - s_t^{c_{i+1}} > \delta_2$ 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 previous 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 4. If $s_t^{c_1} > s_*^{c_1}$, we have that $s_{t+1}^{c_1} \leq s_*^{c_1}$. From 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}.

□

□

This bound is tight; there exist markets in which convergence does not occur in fewer than $N+2$ periods. To see this, let E be a NSW 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": $\eta\{\theta|v^{\theta,c_1} > v^{\theta,c_2} > ... > v^{\theta,c_N}\} = 1-\epsilon$ for some small ϵ and $k_{c_i} < k_{c_j}$ for $0 < i < j$. Moreover, peer preferences are strong: for these $1-\epsilon$ measure of students, $f^\theta(r^\theta, s^{c_i}) > v^{\theta,c_1}$ 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 θ with $r^\theta > s_*^{c_1}$. This follows because 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 8, $s_2^{c_2} < s_*^{c_2}$. 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 $t \leq N$, $s_t^{c_t} < s_*^{c_t}$, which implies that $s_t = s_{N+1}$ only for $t \geq N+1$.

E Theoretical study of non-NSW markets

In this section we investigate non-NSW markets with peer preferences. First, we present an example in which peer preferences cannot be represented via (any finite number of) summary statistics. Second, we study the TIM process in a market where students prefer to attend programs with higher-scoring peers, as this is a functional form of peer preferences studied in Epple and Romano (1998), Rothstein (2006), Beuermann and Jackson (2019), Beuermann et al. (2019), Abdulkadiroğlu et al. (2020), and Avery and Pathak (2021).

E.1 Preferences over the entire distribution of peer ability

Let there be $N \geq 1$ programs. For almost all students θ and all programs $c \in C \setminus \{c_0\}$ we can represent $u^\theta(c|\alpha) = v^{\theta,c} + f^{\theta,c}(\lambda^c(\alpha))$, where

$$f^{\theta,c}(\lambda^c(\alpha)) = -\int_0^1 |\lambda^{c,(1,\ldots,y,\ldots,1)}(\alpha) - \lambda_B^{c,(1,\ldots,y,\ldots,1)}| dy.$$

and where the "y" terms in the superscripts correspond to scores at program c, such that

$$\lambda_B^{c,(1,\ldots,y,\ldots,1)} = \begin{cases} 0 \text{ if } y \leq r^{\theta,c} \\ 1 \text{ if } y > r^{\theta,c} \end{cases}.$$

This functional form represents that each student θ has a "bliss point" and most prefers to attend a program c when her peers at program c all have program c scores equal to $r^{\theta,c}$. For any assignment, the peer cost of attending program c is the difference in area between the actual distribution of types at program c and her bliss point distribution.

Following Definition 4, fix any finite number M of summary statistics $\{s^m(\lambda)\}_{m=1,\ldots,M}$. We claim the peer preferences above cannot be represented via a utility function over $\{s^m(\lambda)\}_{m=1,\ldots,M}$. To see this, fix θ and c, and let θ's utility over program c be characterized as above. First note that the subset of assignments $\hat{\mathcal{A}}$ such that $f^{\theta,c}(\lambda(\alpha)) \neq f^{\theta,c}(\lambda(\alpha'))$ for any $\alpha,\alpha' \in \hat{\mathcal{A}}$ is open and dense in $\mathcal{A}$.[5] Therefore, it suffices to show that there does not exist a function $h^{\theta,c} : [0,1]^M \to \Lambda$ such that $h^{\theta,c}(s^1(\lambda(\alpha)),\ldots,s^M(\lambda(\alpha))) = \lambda^c(\alpha)$ for all α. It is well known that the set Λ has cardinality equal to that of the continuum.[6] Since $h^{\theta,c}(\cdot)$ has only M arguments, it cannot be surjective, implying that there is some α such that $h^{\theta,c}(s^1(\lambda(\alpha)),\ldots,s^M(\lambda(\alpha))) \neq \lambda^c(\alpha)$.

E.2 Ability monotonic preferences

We now consider the case in which students prefer peers with higher ability, in contrast to our modeling of the NSW market. We adopt a common assumption on peer preferences (Epple and Romano, 1998; Avery and Pathak, 2021): student valuation of a program is entirely based on the abilities of peers at that program. We will refer to this as an *ability monotonic preferences market*. We assume there is no entry and exit of programs, and we later argue that entry and exit would not meaningfully change our predictions.

Ability is measured using objective outcomes such as standardized test scores, we assume that student scores are the same for all programs, so that $r^\theta := r^{\theta,c} = r^{\theta,c'}$ for all $c,c' \in C$. For any assignment $\alpha \in \mathcal{A}$, and programs $c,c' \in C \setminus \{c_0\}$ and any student $\theta \in \Theta$, 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 $\theta \in \Theta$.[7] For simplicity of exposition, we assume that for

[5] Any two assignments α,α' that differ among a positive measure set of students will by construction yield $\lambda(\alpha) \neq \lambda(\alpha')$. Recalling that we endow the set Λ with the pointwise convergence topology, the subset of ability distributions $\hat{\Lambda}$ such that $f^{\theta,c}(\lambda) \neq f^{\theta,c}(\lambda')$ for any $\lambda,\lambda' \in \hat{\Lambda}$ is open (Take any $\lambda,\lambda' \in \Lambda$ such that WLOG $f^{\theta,c}(\lambda) = f^{\theta,c}(\lambda') + \delta$ for some $\delta > 0$. There exists sufficiently small ϵ such that $|f^{\theta,c}(\lambda) - f^{\theta,c}(\lambda'')| < \delta$ for any λ'' such that $||\lambda(\cdot) - \lambda''(\cdot)||_\infty < \epsilon$. Therefore, it must be that $f^{\theta,c}(\lambda) \neq f^{\theta,c}(\lambda'')$.) and dense (Fix $\epsilon > 0$. Take any $\lambda,\lambda' \in \Lambda$ such that $f^{\theta,c}(\lambda) = f^{\theta,c}(\lambda')$. It is easy to see that there exists some λ'' such that $||\lambda(\cdot) - \lambda''(\cdot)||_\infty < \epsilon$ such that $f^{\theta,c}(\lambda) \neq f^{\theta,c}(\lambda'')$)

[6] See, for example, Moschovakis (2006, page 18).

[7] Note that any such market does not satisfy Assumption **A2**. As in Example 3, slight adjustments could be made to

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 an ability monotonic preferences market, the TIM process generically generates a stable matching in all time periods $t \geq 1$.[8]

Proposition 5. *Let E be an ability monotonic preferences market. 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 $t = 1$. Without loss of generality let $c_1 \succ^{\theta|\mu_0} c_2 \succ^{\theta|\mu_0} c_3 \succ^{\theta|\mu_0} ... \succ^{\theta|\mu_0} c_N$ for all θ. 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. Moreover, note that $\succeq^{\theta|\mu_0} = \succeq^{\theta|\mu_1}$ for all θ, as the most desired program under μ_0, c_1, remains the most desired program under μ_1 because it enrolls the highest scoring students at μ_1. Similarly, 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, $t \geq 1$.[9]

In contrast to NSW markets, the TIM process converges immediately in an ability monotonic preferences market. An interesting implication is that even with the exit and entry of programs, the TIM process creates a stable matching at every time $t \geq 1$.

F Alternative empirical explanations

F.1 Explanations with "missing mass" prediction

In this section, we describe alternative explanations that predict a "missing mass" of students who top-rank or promote programs with a PYS *just* above their ATAR scores. This is because these alternative explanations all imply a (perceived) cost of top-ranking programs with a non-unity probability of admission. We see no such missing mass, indeed we see that students are more likely to top-rank programs whose PYSs slightly exceed their own ATAR scores (see Figure 2). We similarly do not see a missing mass of students who alter their pre-ROLs to top-rank a program with an ATAR score just above, versus just below, her own ATAR score (see Figure 3).

For these alternative models, we omit time indices and assume that each student θ draws a value $v^{\theta,c} \sim G_c$ independently for each program c, where each G_c:

this market to satisfy **A2**. Our conclusions in this section would not change.

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

[9] 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 throughout the TIM process. Therefore, initiating the TIM process with assignment μ_0' which permutes program identities will lead to a different stable matching.

1. has an associated continuous density g_c, where $g_c(x)$ is positive and bounded away from 0 if and only if $x \in [0,1]$,

2. For any $\epsilon > 0$ there exists $\delta > 0$ such that $||G_c - G_{c'}||_\infty < \epsilon$ if $|PYS_c - PYS_{c'}| < \delta$.

1. is a standard technical assumption generating full support of preferences over programs, and 2. is a continuity condition–students have, in aggregate, similar preferences for programs that have similar observables.

There is some non-increasing function $p(\cdot)$ that maps the difference between a program's PYS and a student's ATAR score into an expected probability of admission, and we take this probability to be independent (conditional on the score gap) across programs. To match our empirical setting, we assume that $p(0) = 1$ and there is a discontinuity at 0–each student perceives a substantially lower probability of admission to a program whose PYS just exceeds her own ATAR score, compared to a program with a PYS just below her ATAR score. This is justified by a non-zero probability of receiving zero bonus points at a program. We denote the magnitude of the discontinuity at 0 by $\Delta > 0$, that is, $\Delta = p(0) - \lim_{x \to 0^+} p(x)$.

F.1.1 Incorrect beliefs

One potential explanation is that students do not fully understand the deferred acceptance mechanism, with a well-known concern being that students do not realize that rejection from a program at the top of their post-ROL does not reduce the probability of matching with a lower-ranked program (Li, 2017). In our context, this implies that students believe there is a cost to top-ranking programs to which they may be rejected (i.e. programs for which the PYS exceeds their ATAR score).

Prima facie evidence does not support this hypothesis; as we discuss in Section III.D, 75% of students top-rank a program on their post-ROL where the program's PYS exceeds the student's ATAR. Below, we formalize this point, and given the salience of the PYS, we argue that under this hypothesis we would expect a "missing mass" of students who top-rank programs on their post-ROLs which have PYSs just above the student's ATAR score.

We assume that each student θ perceives a cost for ranking a program first on her post-ROL and being rejected. We take this cost to be some constant $\kappa > 0$, although our claims apply if we condition this cost on the identity of the program.

Suppose that student θ has an ATAR score of x, and for some small ϵ, consider programs c_1 and c_2, where $PYS_{c_1} \in [x-\epsilon, x]$ and $PYS_{c_2} \in (x, x+\epsilon)$. Because $p(\cdot)$ is non-increasing, it must be that the student perceives at least $1 - p(\epsilon) \geq \Delta$ higher probability of being admitted to program c_1 than c_2. Because of the expected cost of rejection, the student will prefer to rank c_2 rather than c_1 first on her post-ROL only if

$$v^{\theta,c_2} \cdot p(\epsilon) - \kappa(1 - p(\epsilon)) \geq v^{\theta,c_1}$$

Since $1 - p(\epsilon) \geq \Delta$, this condition can only be satisfied if

$$v^{\theta,c_2} - v^{\theta,c_1} \geq [\kappa + v^{\theta,c_2}]\Delta$$

For sufficiently small ϵ the probability that $v^{\theta,c_2} - v^{\theta,c_1} \geq 0$ is approximately $\frac{1}{2}$. Therefore, because $\kappa > 0$ and $\Delta > 0$, θ will be strictly more likely to rank c_2 first on her post-ROL than c_1. Averaging over all students, this implies that the share of students who rank a program first on their post-ROL with a PYS just exceeding their own ATAR score is discontinuously lower than the share of students top-ranking a program with a PYS equalling, or just lower than, their ATAR score.

We see no such discontinuity, indeed we see that students are more likely to top-rank programs whose PYSs slightly exceed their own ATAR scores (see Figure 2). Moreover, we would similarly expect a discontinuously larger fraction of students whose top-ranked program on their pre-ROL has a PYS just exceeding their ATAR score to demote this program compared to students whose top-ranked program on their pre-ROL has a PYS that is just exceeded by their ATAR score. We see no such difference in Figure 3; students with ATAR scores of 86 have an average score gap of zero, and we see similar adjustments to those with ATAR scores just above and below this level.

F.1.2 Non-classical preferences

Non-classical utility functions can also explain some non-standard behavior in matching markets. Dreyfuss, Heffetz, and Rabin (2021) study a model in which students have expectations-based loss aversion. As a result, they may fail to rank otherwise desirable options in strategy-proof mechanisms to avoid disappointment from rejection. Meisner and von Wangenheim (2019) study a related model of loss aversion, and predict that only "top-choice monotone" ROLs–ones in which the student ranks all programs preferred to the top-ranked program in decreasing order of her preferences, and all other programs are ranked in increasing order of her preferences–are optimal under these preferences. Meisner (2021) studies a model where students explicitly dislike rejection from programs.

While this form of preferences may partially explain the ROLs of students in our setting (e.g. a student may fail to rank a desired, out-of-reach program on either ROL), it does not likely explain the margin along which we derive our results; the switching behavior we observe does not generally satisfy top-choice monotonicity.

Moreover, our model with a relatively large number of programs offers a clean test of the data because students have a (discontinuously) higher probability of admission to programs where their ATAR scores exceed the PYS. The loss term κ in Section F.1.1 can represent the loss associated with rejection from Meisner (2021) and a reduced form for other related preferences, such as loss aversion. The fact that $\Delta > 0$ implies that our missing mass prediction in Section F.1.1 hold for these preferences as well.

F.1.3 Optimal information acquisition

One potential explanation is that student preferences change over time. This could possibly be due to exogenous factors (e.g. news coverage of a scandal at a program just prior to submission of the

post-ROL) or strategic choices to acquire information about programs (Hakimov, Kübler, and Pan, 2021; Grenet, He, and Kübler, 2022; Immorlica et al., 2020), where students have incentives not to "waste" information acquisition costs on programs they will be rejected from.

Prima facie evidence does not support the "exogenous factors" hypothesis. 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 exogenous preference changes to rationalize our findings, it would have to be that programs with PYSs closer to a student's eventual ATAR score are systematically receiving a positive "shock" relative to other programs.

We further investigate the potential strategic choice of agents to acquire information about different programs. Formally, suppose that for each student θ and each program c, $v^{\theta,c}$ represents a signal of θ's value for attending program c. Student θ's value for matching with program c is $\hat{v}^{\theta,c} = v^{\theta,c} + \sigma^{\theta,c}$ where $\sigma^{\theta,c} \sim U(-\kappa,\kappa)$ independently across students and programs, for some $\kappa > 0$. Each student θ can privately learn her draw $\hat{v}^{\theta,c}$ for up to one program prior to matching. (Although we assume an "all-or-nothing" information acquisition framework, our conclusions likely extend to many more nuanced frameworks.) If student θ matches to program c, her utility is $U(\hat{v}^{\theta,c})$ where $U(\cdot)$ is a bounded, weakly increasing, and strictly concave function from $[-\kappa,1+\kappa] \to [0,1]$. This captures that students prefer programs for which they have high draws, and the concavity ensures risk aversion.

We make four claims:

First, holding fixed the ROLs of other students, each θ is weakly better off if she learns her value for some program c. In the absence of learning her values for any program, she has a weakly dominant strategy to rank programs in descending order of her signals. Upon learning the value for any program, she will optimally alter this order if and only if the learned utility for the selected program rises or falls below the expected utility from another.

Second, consider two potential signal vectors for student θ, $v^\theta = (v^{\theta,c_0}, v^{\theta,c_1}, v^{\theta,c_1}, ..., v^{\theta,c_N})$ and $\tilde{v}^\theta = (\tilde{v}^{\theta,c_0}, \tilde{v}^{\theta,c_1}, \tilde{v}^{\theta,c_1}, ..., \tilde{v}^{\theta,c_N})$, such that $\tilde{v}^{\theta,c} = v^{\theta,c}$ for all $c \notin \{c_1,c_2\}$, $\tilde{v}^{\theta,c_1} = v^{\theta,c_2}$, and $\tilde{v}^{\theta,c_2} = v^{\theta,c_1}$. That is, $\tilde{v}^\theta$ is obtained from v^θ by permuting the signals of c_1 and c_2. Consider the case in which θ optimally learns $\hat{v}^{\theta,c_2}$ upon receiving signal vector v^θ (we ignore non-generic and non-payoff relevant cases in which there are multiple optimal selections). According to her weakly dominant strategy, θ will rank programs in terms of their expected utility (where her expected utility for c_2 is $U(\hat{v}^{\theta,c_2})$).[10] We claim that θ must then optimally learn $\hat{v}^{\theta,c_1}$ upon receiving signal vector $\tilde{v}^\theta$. Recall that σ^{θ,c_1} and σ^{θ,c_2} are independently and identically distributed, and that θ is guaranteed entry to c_1 but not c_2. Conditional on not matching with a program preferred to c_2 upon observing v^θ and learning $\hat{v}^{\theta,c_2}$, there is a probability of at least $\Delta > 0$ that θ is not admitted to c_2 and the information gathered is therefore payoff irrelevant (the first

[10] Depending on θ's ATAR score, there are payoff equivalent ROLs that omit programs with zero probability of acceptance, or programs that are dispreferred to others which guarantee acceptance. Our conclusions will hold regardless of which of these ROLs is selected.

claim establishes that information acquisition improves expected payoffs). The probability of rejection also implies that θ does not always (i.e. for almost every draw of signals) optimally learn $\hat{v}^{\theta,c_2}$ upon receiving signal vector v^θ if she optimally learns $\hat{v}^{\theta,c_1}$ upon receiving signal vector $\tilde{v}^\theta$.

Third, consider a student θ with an ATAR score of x, and suppose there exist programs c_1 and c_2 such that $PYS_{c_1} \in [x-\epsilon,x]$ and $PYS_{c_2} \in (x,x+\epsilon)$ for $\epsilon > 0$. We claim that for sufficiently small ϵ, θ is, ex-ante, more likely to optimally learn v^{θ,c_1} than v^{θ,c_2}. This follows from the second bullet and the assumption that the signal distributions G_{c_1} and G_{c_2} are arbitrarily close for sufficiently small ϵ and therefore, $\tilde{v}^\theta$ and v^θ are nearly equally likely to occur.

Fourth, student θ with an ATAR score of x is, for sufficiently small $\epsilon > 0$, discontinuously more likely to top-rank program c_1 than program c_2 on her post-ROL if $PYS_{c_1} \in [x-\epsilon,x]$ and $PYS_{c_2} \in (x,x+\epsilon)$. This follows from the third claim, and the fact that $U(\cdot)$ is strictly concave. Therefore, resolution of uncertainty provides student θ an expected utility "boost" from that program.

Averaging over all students, this implies that the share of students who rank a program first on their post-ROL with a PYS just exceeding their own ATAR score is discontinuously lower than the share of students top-ranking a program with a PYS equalling, or just lower than, their ATAR score.

We see no such discontinuity, indeed we see that students are more likely to top-rank programs whose PYSs slightly exceed their own ATAR scores (see Figure 2). Moreover, we would similarly expect a discontinuously larger fraction of students whose top-ranked program on their pre-ROL has a PYS just exceeding their ATAR score to demote this program compared to students whose top-ranked program on their pre-ROL has a PYS that is just exceeded by their ATAR score. We see no such difference in Figure 3; students with ATAR scores of 86 have an average score gap of zero, and we see similar adjustments to those with ATAR scores just above and below this level.

F.2 Mismatch/"fit" preferences

One alternative is that students do not use the PYS to learn about the skill distribution of peers, but rather as an indication of their mismatch or fit with a particular program. They may be uninformed about a program and interpret the PYS as a signal, for example, of how difficult or prestigious the program is for them. Their choices may be influenced by these updates to their program-specific information rather than the peer preference mechanism.

A straightforward prediction is that programs in which students have a stronger prior–perhaps coming from more, or more informative signals excluding the PYS–will be less affected by the signal provided by the PYS.

Empirically, we can implement a test of the "strength" of the signal provided by the PYS, by including program age in our analysis. This assumes that students likely have more information about long-standing programs. Therefore, as remarked in the previous paragraph, the PYS provides relatively less information about older, established programs.

We test this hypothesis for both of our identification strategies. In our across-person analysis, we

interact PYS with program age in Equation 1. Under the hypothesis of "fit" preferences, the effect of the PYS on student demand should dissipate for older programs.

Table A.1 shows the impact of the PYS on student 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:

$$\begin{aligned} y_{c,t} = &\beta PYS_{c,t} + \gamma Age_{c,t} + \lambda Age\ Known_c + \delta_0 PYS_{c,t} \times Age_{c,t} + \delta_1 PYS_{c,t} \times Age\ Known_c \\ &+ \delta_2 Age_{c,t} \times Age\ Known_c + \delta_3 PYS_{c,t} \times Age_{c,t} \times Age\ Known_c + \alpha_c + \alpha_t + \epsilon_{c,t} \end{aligned} \tag{A.10}$$

where $y_{c,t}$ denotes the average applicant score, the number of students who apply, the 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.1 presents the linear combination of coefficients of Equation A.10 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 in existence for 14 or more years ($Age\ Known_c$=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 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 in Figure 5, the oldest programs typically have the highest PYS values, 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.

We now turn to a similar test in our within-person analysis. Under this hypothesis, we similarly predict that the effect of the PYS/ATAR score gap on student demand should dissipate for older programs. We estimate impact of the PYS/ATAR score gap on student demand for very young (2 years old) versus very old programs (7 or more years old) in our sample. Specifically, we estimate the following regression:

$$\begin{aligned} y_{c,t,\theta} = &\beta(PYS_{c,t} - ATAR_i) + \gamma Age_{c,t} + \lambda Age\ Known_c + \delta_0 (PYS_{c,t} - ATAR_i) \times Age_{c,t} \\ &+ \delta_1 (PYS_{c,t} - ATAR_i) \times Age\ Known_c + \delta_2 Age_{c,t} \times Age\ Known_c \\ &+ \delta_3 (PYS_{c,t} - ATAR_i) \times Age_{c,t} \times Age\ Known_c + \epsilon_{c,t,\theta} \end{aligned} \tag{A.11}$$

where $y_{c,t}$ denotes an indicator for "promote." $Age_{c,t}$ 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. Table A.2 presents the linear combination of coefficients of Equation A.11 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 in existence for 7 or more years ($Age\ Known_c = 0$).

We note that the coefficient on our outcome of interest ("promote") in all columns is larger and statistically significant for the oldest programs, which we would not expect if students cared about their fit with a program and not their peers.

G Additional evidence, figures, and tables

Figure A.1 visually presents our two identification strategies and the assumptions required for each.

Across Person analysis

Table A.3 studies student preference for program "quality" in a regression framework. The dependent variable is the number of times a program is ranked by student post-ROLs in a given year, while the main regressor is the program PYS. We include field of study and year fixed effects to isolate cross-sectional variation. We find that programs with higher PYSs are more likely to be ranked by students. This suggests an aggregate preference for higher-quality programs amongst students.

Our test in Table 2 assumes that students form rankings based on the PYS due to peers present in programs, and not due to changes in program quality reflected by the PYS. To test this assumption, we include lagged values of a programs' PYS in Equation 1. Table A.4 presents results. Lagged values of the PYS (from two and three years before the current year) have little predictive power, indicating that responses to the PYS are not based on a trend of changes in the PYS over time.

Within Person analysis

Table A.5 presents summary statistics on the adjustments students make between their pre- and post-ROLs.

We provide further evidence that changes to ROLs reflect big-fish preferences. We estimate Equation 2 where we replace the left hand side with indicators for whether program c was added or removed by student θ in year t on the PYS/ATAR score gap between c and θ. Following our big-fish hypothesis, programs that are added within the ROL have PYSs that are systematically lower than those that are removed, and are closer to the student's ATAR score. We note that the margin for removals is weaker.

Changes in the NSW market over time, and additional convergence results

Figure A.2 displays how student changes over programs (which may exit and enter) vary over time, while student preferences over universities and field of study are relatively constant across time. This

provides support for our theoretical modeling that the differences across years in the market are due to changes in the set of available programs.

Figure A.3 recreates the plot of average change in PYS from Figure 7 (top panel) and controls for entry and exit of programs into our data by grouping programs by the number of years each program is observed in our data (bottom panel). Across all groups, we observe a similar decreasing trend in the absolute change in PYS over time, which falls to under 1 point as programs age beyond 10 years (four of five of the groups in our data for at least 10 years have all point estimates beyond year 10 less than 1 point.) This provides evidence that the trend we observe in Figure 7 is not due to composition changes over time.

Figure A.4 investigates how the PYSs of programs with similar PYSs in 2012 (left panel) and 2016 (right panel) evolve over time. We plot the average PYS for this group in each year in our sample. We observe that this average PYS converges over time in both panels.

Additional results on program attrition

We find evidence that 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.[11] We test for a significant relationship between yearly changes in PYS and the share of low-SES students in Table A.8. This relationship is positive and significant; a 1 point increase in CYS-PYS is associated with a 2% increase in the percent of low SES students attending the program. This pattern is robust to controls for year, field of study, and program age. We find a similar pattern when looking at the share of minority, disabled, and rural-based students across universities with more- or less-volatile PYSs (see Figure A.5).

These results suggest that programs with PYSs not in steady state 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, the results do show that the population most impacted by nonconvergence has a lower socioeconomic status, and includes those who are less likely to complete their studies at their initial program.

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

Figure A.1: Theory to Data – Necessary Assumptions and Tests to Identify Peer Preferences

If:

Untestable assumption	Strong version: ROL provides truthful revelation of **ordinal** preferences, given the PYS
Features that support untestable assumption	・DA mechanism used is strategy proof for those with < 9 acceptable programs ・Acceptance probability ∈ (0,1) ・When creating ROL, students are only shown PYS
Additional conditions on data	・Restrict to students with ROL length < 9
Remaining Caveats	・ROL is also determined by probability of acceptance, or cost to listing "reach" programs

Then verify the following:

Testable assumptions	1) Student ROL is responsive to program PYS	2) The PYS teaches about peers, not only program characteristics or trends
Empirical Test	・Do changes in program PYS lead to changes in ROL? ・See Eq. 1	・Add interaction of program age with PYS to Eq. 1 ・Add lagged PYS to RHS of Eq. 1
Outcome in our data	・Yes -- a higher PYS causes fewer low-scoring students to list the program	・Effect of PYS not significantly smaller for older, established programs ・Lagged PYS coefficients small and insignificant

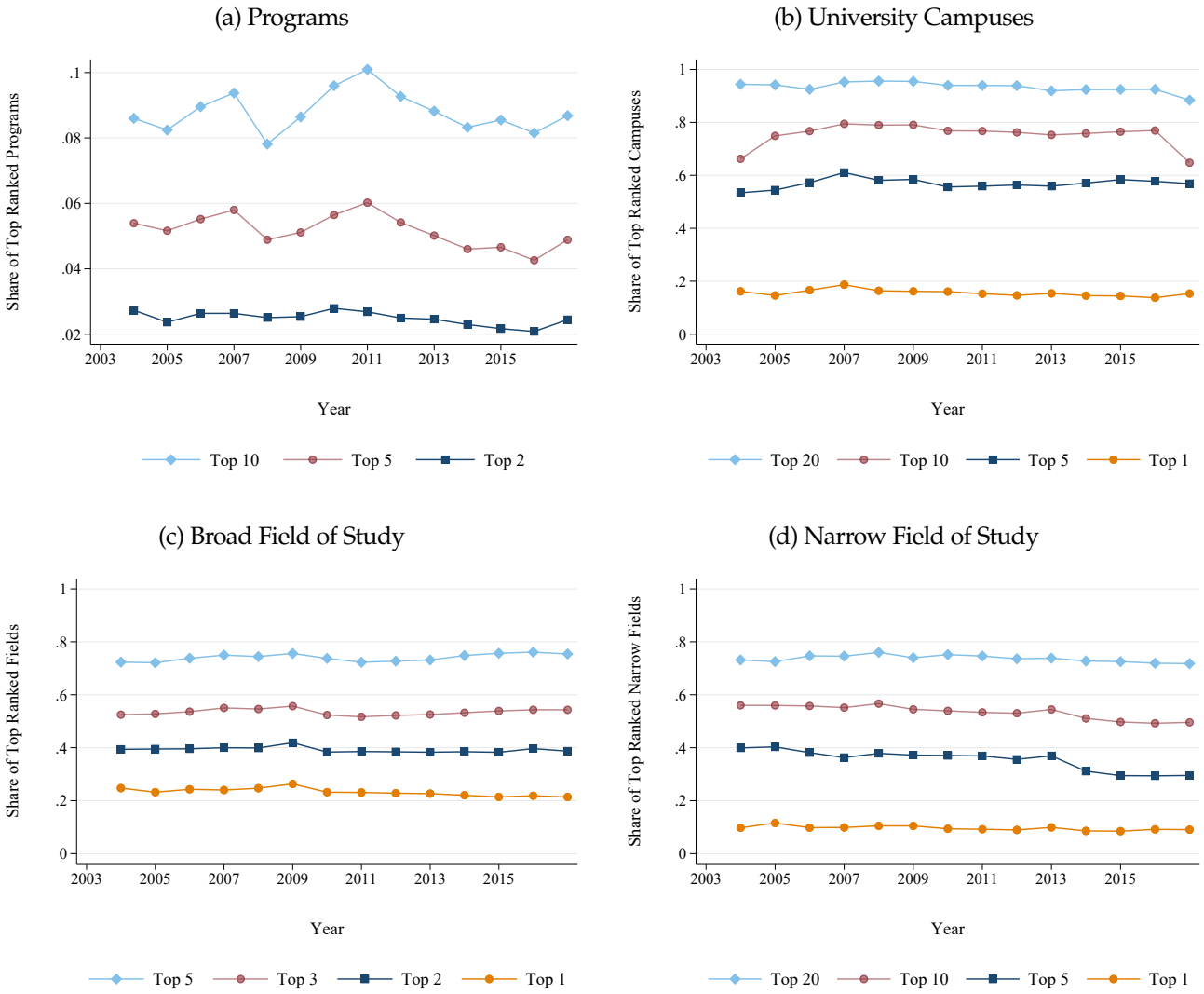
If:

Untestable assumption	Weak version: ROL provides truthful revelation of **relative** preferences, given the PYS
Features that support untestable assumption	・Weakly dominated to submit wrong relative order in ROL ・Acceptance probability ∈ (0,1) ・When creating ROL, students are only shown PYS
Additional conditions on data	・Use 2 ROLs per person, one before one after ・Restrict analysis to "switches"

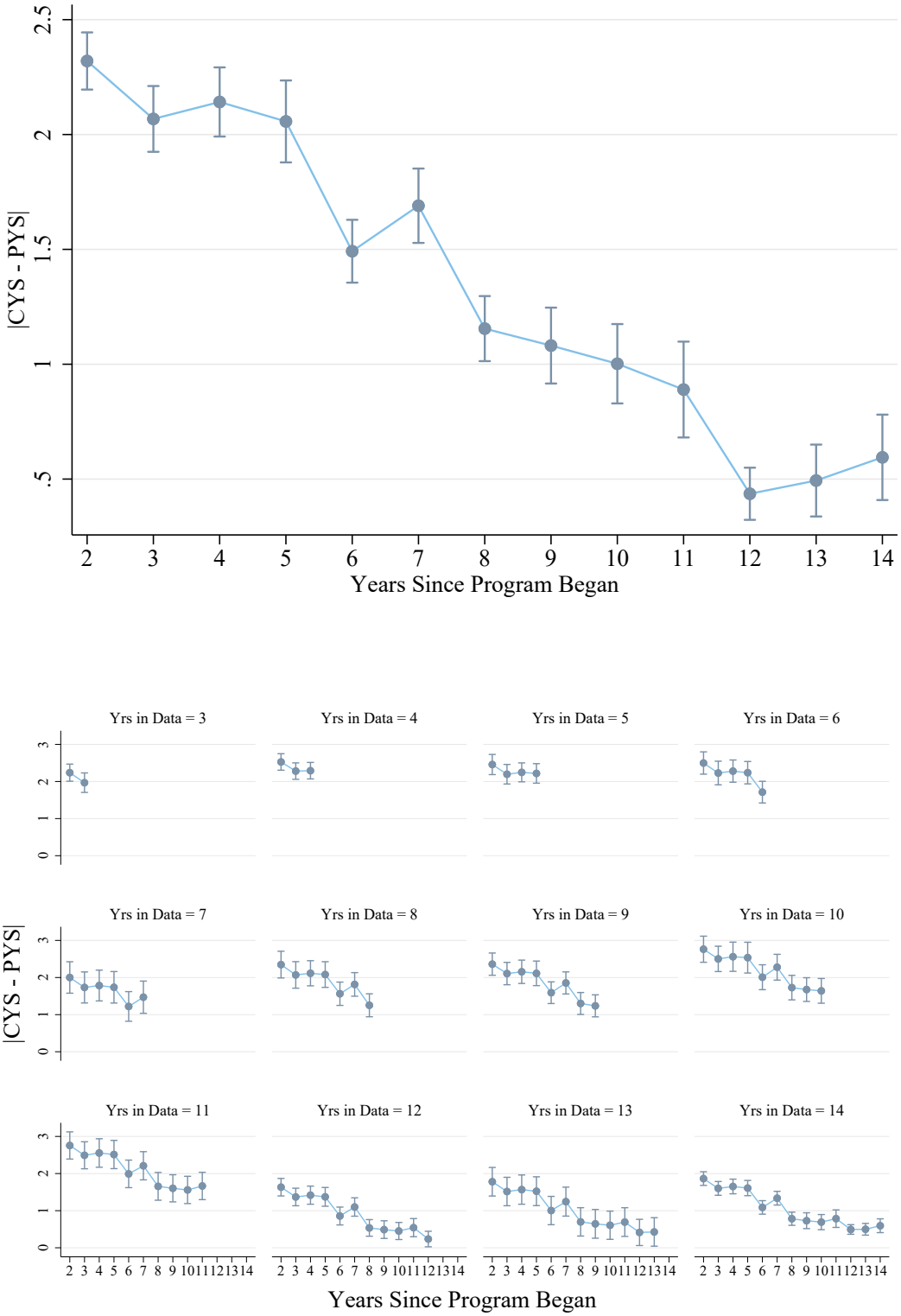
Then verify the following:

Testable assumptions	1) Student ROL is responsive to program PYS	2) The PYS teaches about peers, not only program characteristics or trends
Empirical Test	・Do "switches" in ROL correlate with the PYS/ATAR score gap? ・See Eq. 2	・Add interaction of program age with PYS to Eq. 2
Outcome in our data	・Yes -- students are more likely to promote programs with a smaller score gap.	・Effect of PYS not significantly smaller for older, established programs
Testable assumptions	3) pre-ROL contains meaningful information on preferences, changes do not reflect "random" changes in preferences	4) Changes between two ROLs outcome relevant
Empirical Test	・Measure correlation between pre- and post-ROL ・Test whether changes to ROL are predicted by (initially unknown) student ATAR score	・Measure what percentage of switches lead to a difference in ultimate match.
Outcome in our data	・Strong overlap between two ROLs ・Changes to ROL predicted by score gap	・54% of students who change ROL receive different matching -- See Section III.D.3

Figure A.2: Aggregate student preferences over programs, campuses, and fields

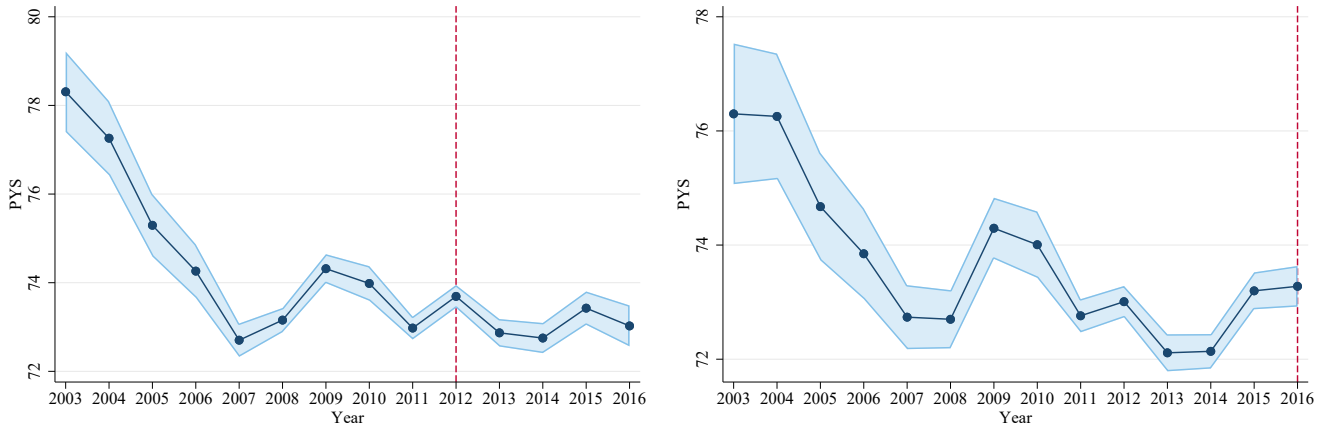


This figure shows that student preferences for program, campus, and field of study are broadly stable across cohorts. We "fix" a group of campuses or programs based on overall popularity, and then show that this popularity ranking is stable year to year. To create the campus graph, for example, we use the entire panel dataset of applications. We count how many times each campus was ranked either first or second on a student'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 a student'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 most popular overall campus, consistently receives the top-ranking for about 20% of students 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.



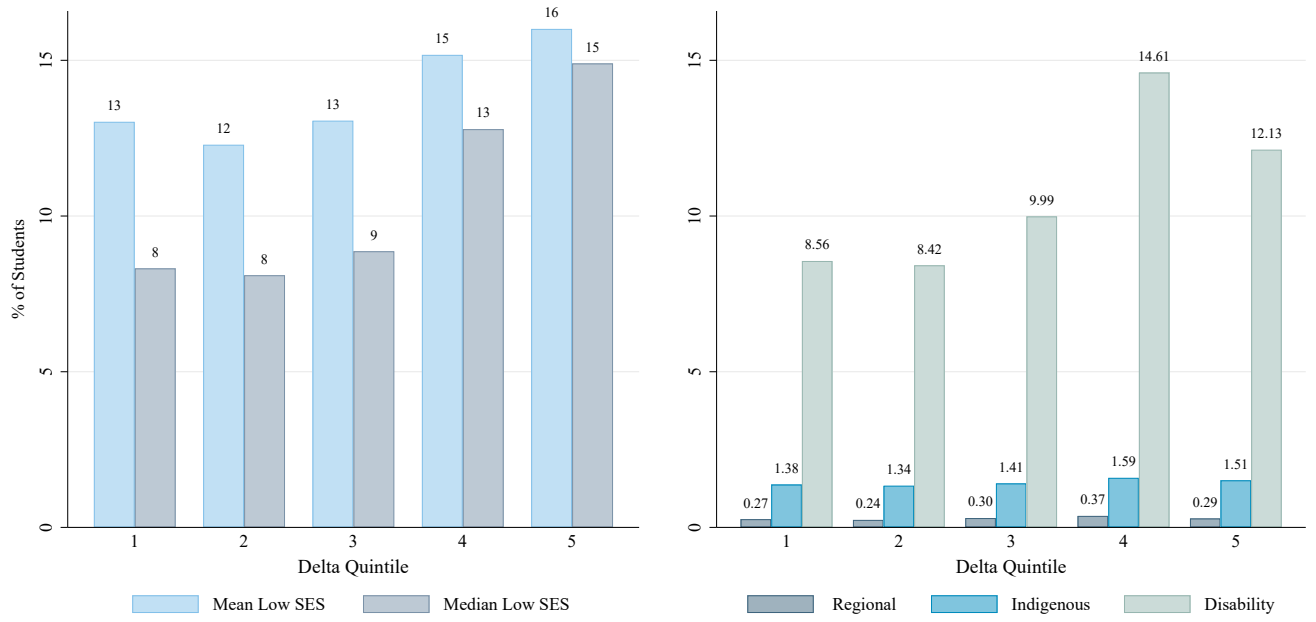
These figures provide evidence for convergence over time in program PYSs. We plot $\Delta_{c,t} := |CYS_{c,t} - PYS_{c,t}|$, a measure of PYS "instability" against the number of years the program has existed. They show that the PYS converges with time (top figure), and that this panel is not driven by the entry or exit of programs into the sample (bottom figure). 95% confidence intervals using standard errors clustered at the program level are indicated.

Figure A.4: Convergence test for matching in 2012 and 2016



The left panel groups together programs that have a similar PYS (within a 10-point band of 70) in 2012, and follows the group's **average** PYS both forward and backward in time. It shows that programs with similar PYSs in 2012 have converged from a more dispersed distribution over time, and continue to converge even after 2012. The right figure repeats the same exercise, instead grouping together programs with a similar PYS in 2016. 95% confidence intervals using standard errors clustered at the program level are indicated.

Figure A.5: Demographics and CYS-PYS



We test whether the instability generated by disregarding peer preferences is borne primarily by students from less advantaged demographic backgrounds. We merge in data on gender, ethnicity, and socioeconomic status at the university-year level. Following Remark 7, we focus on program-years with CYS>PYS. We divide our sample of university programs into quintiles, based on the size of their yearly changes in PYS (i.e. higher quintiles correspond to larger values of CYS-PYS. We find that programs in the highest quintiles are at universities with a larger share of low-SES students. These universities also tend to serve more students with disabilities, those from rural areas, and those with indigenous backgrounds. While these results do not claim to show causality in either direction, they show that the population most impacted by non-convergence has a lower socioeconomic status.

Table A.1: Impact of PYS on Student 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	0.282***	-1.947***	-0.007***	-0.003	-0.013***
	(0.023)	(0.195)	(0.001)	(0.002)	(0.001)
14+ Year Old Programs	0.359***	-1.711***	-0.007***	-0.039*	-0.014***
	(0.029)	(0.546)	(0.002)	(0.022)	(0.003)
Observations	14,853	14,853	14,853	14,853	14,853

This table shows the linear combination of estimated coefficients for Equation A.10 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. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A.2: Impact of Score Gap on Promote for New and Established Programs

	(1) Promote	(2) Promote	(3) Promote	(4) Promote	(5) Promote	(6) Promote
2 Year Old Programs	0.00084	0.00068	0.00085	0.00092	0.00109	0.00127*
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
7+Year Old Programs	-0.00213***	-0.00188***	-0.00217***	-0.00208***	-0.00196***	-0.00178***
	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Program FE		✓				
ROL length FE			✓			
Top Progam FE				✓		
Top 2 Programs FE					✓	
Top 3 Programs FE						✓
Observations	537,442	537,406	537,442	537,442	537,442	537,442

This table shows the linear combination of estimated coefficients for Equation A.11 for (i) 2 year old programs (ii) 7 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 7 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 7 or more years old (*Age Known*$_c$=0). Column (2) includes program fixed effects, column (3) includes a fixed effect for the number of programs listed on a student's pre-ROL, column (4) includes a fixed effect for the top-ranked program in the pre-list, column (5) includes a fixed effect for the top two ranked programs in the pre-list, column (6) includes a fixed effect for the top three ranked programs in the pre-list. We use the pre- and post-ROL sample from 2010-2016. Standard errors in parentheses, clustered at the program level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A.3: Relationship between program PYS and popularity amongst students

	Program listed on ROL			
	(1)	(2)	(3)	(4)
PYS	0.81**	0.87**	0.96**	1.02**
	(0.28)	(0.29)	(0.31)	(0.32)
Year FE		✓		✓
Field FE			✓	✓
	Program listed first on ROL			
	(1)	(2)	(3)	(4)
PYS	0.31***	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 a student'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 students generally prefer to rank programs with higher PYSs. We refer to this as a preference for program quality. Standard errors in parentheses, clustered at program level. $^{*}p<0.05$,$^{**}p<0.01$,$^{***}p<0.001$

Table A.4: Across Time Student Response to Program PYS, including Lagged PYS Values

	(1) Avg. Applicant Score	(2) # of Applicants	(3) % of Applicants	(4) % of Applicants Higher Score	(5) % of Applicants Lower Score
Previous Year's Statistic	0.292***	-2.560***	-0.010***	-0.007***	-0.017***
	(0.019)	(0.284)	(0.001)	(0.002)	(0.002)
2 Years Ago Statistic	0.038*	-0.137	0.000	0.003*	0.001
	(0.016)	(0.189)	(0.001)	(0.001)	(0.001)
3 Years Ago Statistic	0.075***	-0.528*	0.001	0.003**	-0.002
	(0.015)	(0.231)	(0.001)	(0.001)	(0.001)
Observations	8,680	8,680	8,680	8,680	8,680

This table shows the estimated β coefficients of a regression similar to (1) where we additionally include the 2 and 3 Years Ago Statistic of the program. $y_{c,t}$ is the average applicant score, the number of students who apply, the 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.5: Summary Statistics of Adjustments to pre-ROL

Variable	Obs	Mean	Std. Dev.	P25	P50	P75
Only Switchers	104,519	0.13	0.33			
Only Adders	104,519	0.07	0.26			
Only Removers	104,519	0.03	0.16			
Any Switch	104,519	0.24	0.43			
Any Add	104,519	0.29	0.45			
Any Remove	104,519	0.21	0.4			
Any Change	104,519	0.46	0.5			
Nr. of Switches	104,519	.77	1.87	0	0	1
Nr. of Adds	104,519	.67	1.36	0	0	1
Nr. of Removes	104,519	.49	1.19	0	0	0
Nr. of Changes	104,519	1.93	2.91	0	0	3
Share of final list Switched	104,519	0.14	0.25	0	0	0.25
Share of final list Added	104,519	0.09	0.18	0	0	0.13
Share of final list Removed	104,519	0.06	0.13	0	0	0
Share of final list Changed	104,519	0.29	0.36	0	0	0.63

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.6: Impact of Score Gap on Add

	(1) Add	(2) Add	(3) Add	(4) Add	(5) Add	(6) Add
PYS - ATAR	-0.0018***	-0.0015***	-0.0019***	-0.0015***	-0.0013***	-0.0010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Program FE		✓				
ROL length FE			✓			
Top Progam FE				✓		
Top 2 Programs FE					✓	
Top 3 Programs FE						✓
Observations	537,442	537,406	537,442	537,442	537,442	537,442

The dependent variable is an indicator for whether a program was added to a student's post-ROL. Column (2) includes program fixed effects, column (3) includes a fixed effect for the number of programs listed on a student's pre-ROL, column (4) includes a fixed effect for the top-ranked program in the pre-list, column (5) includes a fixed effect for the top two ranked programs in the pre-list, column (6) includes a fixed effect for the top three ranked programs in the pre-list. We use the pre- and post-ROL sample from 2010-2016. Standard errors in parentheses, clustered at the program level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A.7: Impact of Score Gap on Remove

	(1) Remove	(2) Remove	(3) Remove	(4) Remove	(5) Remove	(6) Remove
PYS - ATAR	0.0001	0.0000	0.0001*	0.0001*	0.0001*	0.0001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Program FE		✓				
ROL length FE			✓			
Top Progam FE				✓		
Top 2 Programs FE					✓	
Top 3 Programs FE						✓
Observations	537,442	537,406	537,442	537,442	537,442	537,442

The dependent variable is an indicator for whether a program was removed from a student's pre-ROL. Column (2) includes program fixed effects, column (3) includes a fixed effect for the number of programs listed on a student's pre-ROL, column (4) includes a fixed effect for the top-ranked program in the pre-list, column (5) includes a fixed effect for the top two ranked programs in the pre-list, column (6) includes a fixed effect for the top three ranked programs in the pre-list. We use the pre- and post-ROL sample from 2010-2016. Standard errors in parentheses, clustered at the program level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A.8: Relationship Between Percent of Low SES Students and CYS-PYS

Program-years with CYS-PYS > 0 (*N = 4,094*)			
CYS-PYS	0.368***	0.340***	0.272***
	(0.042)	(0.044)	(0.043)
Means: Low SES Percent = 15.43; CYS-PYS = 2.5			
Year FE	✓	✓	✓
Field FE		✓	✓
Course Age FE			✓

This table tests for the relationship between the year-to-year change in PYS of a given program and the percent of its students with a low socioeconomic background. Following Remark 7, we focus on program-years with CYS>PYS. We find that, across a host models with various fixed effects, programs with more volatility in their PYS also have higher share of low SES students. Standard errors in parentheses, clustered at the program level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.