Beyond Truth-telling: A Replication Study on School Choice

Andersson, Tommy; Kessel, Dany; Lager, Nils; Olme, Elisabet ; Reese, Simon

2024

Document Version:
Other version

Link to publication

Citation for published version (APA):
Andersson, T., Kessel, D., Lager, N., Olme, E., & Reese, S. (2024). *Beyond Truth-telling: A Replication Study on School Choice*. (pp. 1-13). (Working Papers; No. 2024:1).

Total number of authors:
5

Working Paper 2024:1

Department of Economics
School of Economics and Management

Beyond Truth-telling: A Replication Study on School Choice

Tommy Andersson
Dany Kessel
Nils Lager
Elisabet Olme
Simone Reese

January 2024
Revised: Oct 2024

LUND UNIVERSITY

Beyond Truth-telling:
A Replication Study on School Choice*

Tommy Andersson,† Dany Kessel,‡ Nils Lager,§ Elisabet Olme,¶ and Simon Reese‖

October 24, 2024

Abstract

In a recent paper, Fack et al. (2019, American Economic Review) convincingly argue and theoretically demonstrate that there may be strong incentives for students to play non-truth-telling strategies when reporting preferences over schools, even when the celebrated deferred acceptance algorithm is employed. Their statistical test also rejects the (weak) truth-telling assumption in favour of another assumption, called stability, using a single data set on school choice in Paris. This paper uses Swedish school choice data and replicates their empirical finding in 66 of the 75 investigated data sets (P-value threshold 0.05).

Keywords: school choice, deferred acceptance algorithm, truth-telling, stability, replication study.

JEL Classification: D12, D82, I23.

*We are grateful to Julien Grenet for valuable comments on this paper, help with the replication package and for taking time to answer our follow-up questions. Andersson, Kessel, Lager, and Olme gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation (Andersson and Lager, grant no. P22-0087; Kessel, grant no. W18–0004; Olme, Browaldh stipend). The Swedish Ethical Review Authority has approved this study (decision no. "Dnr 2022–04825–01").

†Department of Economics, Lund University and Stockholm School of Economics (corresponding author). E-mail: tommy.andersson@nek.lu.se.

‡Department of Economics, Södertörn University.

§Department of Economics, Stockholm School of Economics.

¶Svensk Utbildningsteknologi AB, Stockholm.

‖Department of Economics, Lund University.

1 Introduction

School districts in Chile, France, India, Sweden, the United States and many other countries use the deferred acceptance algorithm (Gale and Shapley, 1962) to assign students to schools.[1] When this algorithm is employed and student preferences and school priorities are strict (possibly after applying some exogenous tie-breaking rules), all students are assigned to their most preferred reported school among the ones that respect all school priorities (Gale and Shapley, 1962) and truthful preference revelation is a dominant strategy for the students (Abdulkadiroğlu and Sönmez, 2003; Roth, 1982).

But even if the deferred acceptance algorithm is strategy-proof in an ideal world, there are reasons to believe that students often deviate from the strict truth-telling strategy, i.e., they do not submit a truthful ranked-ordered list (ROL) containing *all* schools in the district.[2] If the district contains many schools, students rarely rank *all* of them, for example, because they find it cognitively costly or time-consuming (even if a limited number of schools are ranked in accordance with the true preferences, the strategy is only weakly truth-telling since not *all* schools are included in the ranked-ordered list). Another reason is that school districts often only allow students to rank a limited number of schools and students may therefore decide to strategically rank "safe schools," i.e., schools where their probability of admission is close to one, thus deviating from the strict truth-telling strategy (see Haeringer and Klijn, 2009; Romero-Medina, 1998). Furthermore, as observed by Fack et al. (2019), truth-telling is only a *weakly* dominant strategy when the deferred acceptance algorithm is employed, leaving open the issue of multiple equilibria. In other words, there may be strategies that are not dominated by the truth-telling strategy, for example, the strategy not to rank "impossible schools" for which the expected probability of admission, based on available information such as previous admission outcomes, is zero. As theoretically demonstrated by Fack et al. (2019), if both student preferences and school priorities are private information and if there is a cost associated with the cognitive task to submit a rank-ordered list, the strategy of "skipping the impossible" schools may even dominate the

[1] See Abdulkadiroğlu and Andersson (2023) for a recent survey on the theory and practice of school choice.

[2] Students often have incentives to be strategic, rather than truthful, in school choice environments. Strategic behaviour has been observed in school choice programs all over the world (see, e.g., Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2005b,a; Agarwal and Somaini, 2018; Burgess et al., 2015; de Haan et al., 2023; Dur et al., 2018; Fack et al., 2019; Hastings et al., 2009; He, 2017) as well as in laboratory experiments (see, e.g., Chen and Sönmez, 2006; Hakimov and Kübler, 2021). See Rees-Jones and Shorrer (2023) for a recent review.

truth-telling strategy.

It is important to investigate which strategies that students use, for example, because when students rationally play non-truth-telling strategies, other econometric techniques are required to reveal the underlying preferences, compared to the standard case when the researchers can trust preference rankings to be weakly truth-telling. Fack et al. (2019) convincingly argue (using some of the above arguments) and theoretically demonstrate that truth-telling may be a too strong assumption to impose on student preferences, even if the deferred acceptance algorithm is employed, and test an empirical model of school choice based on a truth-telling assumption against one which relies on a weaker assumption referred to as *stability*. Under the stability assumption, the schools in the ranked-ordered lists are preferred in the stated order, but students are only assumed to prefer schools in the list to ex post feasible schools that not are included in the list, i.e., schools that the student would have been admitted to if she had chosen to include them in the list. Note that the stability assumption, in contrast to the weak truth-telling assumption, does not reveal if a school that is included in the ranked-ordered list is preferred to a non-ranked and non-feasible school. Consequently, stability, but not weak truth-telling, allows students to "skip the impossible."

Fack et al. (2019) compare the weak truth-telling and stability assumptions for a single data set on Paris school choice data, containing 1,590 students, and reject the null hypothesis of weak truth-telling. More precisely, a Hausman test on the parameters of models based on weak truth-telling and stability assumptions, respectively, reveals significant differences between the estimated utilities of admission to a particular school or the change in utility of admission to the nearest possible school. The method developed by Fack et al. (2019) has been adopted in many recent paper, including, e.g., Che et al. (2023), Combe et al. (2022a,b), Hahm and Park (2022), and Otero et al. (2021).

While most empirical studies on school choice, like Fack et al. (2019), have been conducted on at most a handful of data sets (see, e.g., Hastings and Weinstein, 2008; Abdulkadiroğlu et al., 2017, 2020; Beuermann et al., 2022; de Haan et al., 2023), this paper uses data from multiple years and grades from 15 Swedish school districts, resulting in 75 different data sets with an average of 887 students. The admissions in all of them are based on the deferred acceptance algorithm, thus making them ideal to investigate if the empirical results in Fack et al. (2019) can be replicated or not.

We report adjusted P-values, corrected for multiple hypothesis testing using the sequential method of Holm (1979). We find that the null hypothesis of weak truth-telling is rejected in 66 of the 75 investigated data sets (P-value threshold 0.05), thus providing strong support for the finding in Fack et al. (2019).

The remaining part of this paper is organized as follows. Section 2 introduces the school choice model and some basic definitions, including formal descriptions of student strategies. The empirical approach is described in Section 3. The data and the results can be found in Section 4.

2 The School Choice Model

2.1 Preliminaries

Students and schools are collected in the finite sets $\mathcal{I} = \{1, \dots, n\}$ and $\mathcal{S} = \{1, \dots, m\}$, respectively. Each student $i \in \mathcal{I}$ has strict preferences over the schools in $\mathcal{S}$ represented by a binary and complete preference relation $\succ_i$.[3] Here, $s \succ_i s'$ means that student i strictly prefers school s over school s'. The preferences of student i will, for convenience, often be represented by a utility score for each school. Formally, a utility vector is a vector of von Neumann-Morgenstern utilities $u_i = (u_{i1}, \dots, u_{im}) \in \mathbb{R}^m$ such that u_{is} is a real number describing the utility that student i obtains by being admitted to school s. Let $\succ$ and u represent the collection of the preferences and the corresponding utility representation, respectively, i.e., $\succ = (\succ_1, \dots, \succ_n)$ and $u = (u_1, \dots, u_n)$.

Students are asked by their school districts to report a ranked-ordered list of the schools in the district, but this does not necessarily mean that the students have to rank *all* schools in $\mathcal{S}$. Thus, a ranked-order list for student i is a represented by a list $L_i = (L_{i1}, \dots, L_{ik})$ for some $k \leq m$, where L_{i1} represents the school that student i has reported as the most preferred, L_{i2} represents the school that student i has reported as the second most preferred, and so on. Note that the rank-ordered lists need not represent the true preferences, they only represent the preferences that the students report to the school district (see the discussion in Section 2.2). Let $\mathcal{L}$ denote the

[3] For simplicity and without loss of generality, it is assumed that all students prefer any school in $\mathcal{S}$ to being unassigned. This is also consistent with the Swedish legislation and our empirical strategy since schooling is mandated by law for all students in our data sets.

collection of the ranked-ordered lists, i.e., $\mathcal{L} = (L_1, \ldots, L_n)$.

Each school $s \in \mathcal{S}$ has priorities over the students in $\mathcal{I}$, represented by $\pi_s = (\pi_{1s}, \ldots, \pi_{ns})$. More specifically, student i is assigned an individual priority $\pi_{is} \in [0, 1]$ by school s, where $\pi_{is} < \pi_{js}$ means that student i has a higher priority than student j at school s.[4] Let π represent the collection of school priorities, i.e., $\pi = (\pi_1, \ldots, \pi_m)$. For convenience, the vector $\pi_i = (\pi_{i1} \ldots, \pi_{im})$ will sometimes be used to describe the priority of student i for each school in $\mathcal{S}$. The capacities of the schools are gathered in the vector $\mathcal{Q} = \{q_1, \ldots, q_m\}$.

A matching μ is a function such that each student $i \in \mathcal{I}$ is assigned a school $s \in \mathcal{S}$, i.e., $\mu(i) = s$, and each school $s \in \mathcal{S}$ is matched with a set of students up to its capacity, i.e., $\mu(s) \subset \mathcal{I}$ and $|\mu(s)| \leq q_s$, and $\mu(i) = s$ if and only if $i \in \mu(s)$. The inverse function $\mu^{-1}(s)$ describes which students that have been assigned to school s, i.e., $\mu^{-1}(s) \equiv \{i \in \mathcal{I} : \mu(i) = s\}$.

A matching μ is stable if there does not exist any student–school pair $(i, s) \in \mathcal{I} \times \mathcal{S}$ such that student i prefers school s to her assigned school, i.e., $s \succ_i \mu(i)$, and either there are available seats at school s, i.e., $|\mu(s)| < q_s$ or there is a student i' that is assigned to school s and ranked lower than student i by school s, i.e., $\mu(i') = s$ and $\pi_{is} < \pi_{i's}$.

The realised cutoff $c_s(\pi, \mu)$ at school s at a given matching μ equals 1 if the school is under-subscribed, and otherwise the priority of the "last admitted" student, that is:

$$c_s(\pi, \mu) = \begin{cases} 1 & \text{if } |\mu^{-1}(s)| < q_s \\ \max\{\pi_{is} : i \in \mu^{-1}(s)\} & \text{otherwise} \end{cases}$$

School s is (ex post) feasible for student i if the student's priority π_{is} to school s is below the school-specific cutoff $c_s(\cdot)$. Finally, a matching mechanism determines a matching for any given problem $(\mathcal{I}, \mathcal{S}, \mathcal{Q}, \mathcal{L}, \pi)$.

2.2 Student Strategies

The rank-ordered list L_i is a reported ordinal ranking of schools, while $\succ_i$ represents the true preferences. Because students need not report complete, or even true, rankings, it is important to distinguish between different strategies. To do that, it is first important to define a student type.

[4] School priorities may be constructed, e.g., based on walk zones or sibling priorities. Schools may then have to use a tie-breaker to make the priorities strict.

The type of student i is a description of her utility profile and school priorities, i.e., a pair $\theta_i = (u_i, \pi_i)$. When students are asked to report their ranked-ordered lists to the school districts, they will base their strategy on their type. Formally, a strategy is a function $\phi_i : \mathbb{R}^m \times [0,1]^m \rightarrow L_i$ that maps the type of student i to some rank-ordered list $L_i = (L_{i1}, \ldots, L_{ik})$. Note, in particular, that $k \leq m$, meaning that the strategy for the student need not include all schools in $\mathcal{S}$.

Based on the above definitions, it is possible to define several different strategies. As already explained, there are reasons to believe that students not rank all schools in the school district, especially if it contains many schools. Furthermore, if a student believes she will be admitted to one of her, say, three most preferred schools with probability one, the student is indifferent between stating and leaving out all other schools in her rank-ordered list. Such strategy is referred to as weak truth-telling (WTT).

Definition 1. The strategy $\phi_i^{\text{WTT}}(\theta_i) = L_i^{\text{WTT}}(\theta_i) = (L_{i1}^{\text{WTT}}(\theta_i), \ldots, L_{ik}^{\text{WTT}}(\theta_i))$ for some $k \leq m$ where student i ranks all schools in the list $L_i^{\text{WTT}}(\theta_i)$ in accordance with her true preferences $\succ_i$ is called a weak truth-telling strategy. If $k = m$, the strategy $\phi_i^{\text{STT}}(\theta_i) = L_i^{\text{STT}}(\theta_i) = (L_{i1}^{\text{STT}}(\theta_i), \ldots, L_{im}^{\text{STT}}(\theta_i))$ is called a strict truth-telling strategy.

The difference between strict truth-telling and weak truth-telling is that the latter strategy allows students to omit irrelevant schools at the bottom of their utility profiles.[5] Fack et al. (2019) introduce an application cost for students when ranking multiple schools and show (in their Proposition 1) that the STT strategy is the unique Bayesian Nash equilibrium under the deferred acceptance algorithm if there are no application costs for any student and if the joint distribution of preferences and priorities has full support. However, for any non-zero application costs, there always exist student types for which the STT strategy is not an equilibrium strategy.

Fack et al. (2019) also clarify the relationship between weak truth-telling and stability. More precisely, they show (in their Proposition 3) that if every student plays the WTT strategy under the deferred acceptance algorithm (which may not be an equilibrium), then given a realized matching (i) whenever a student is assigned to a school, she is matched to her favorite feasible school and (ii) if all students that have at least one feasible school are assigned to some school, the matching is stable. This finding has implications for their empirical approach.

[5] It has become a commonly used assumption in the empirical school choice literature that students are at least weak truth-telling when the deferred acceptance algorithm is applied, see, e.g., Abdulkadiroğlu et al. (2017).

3 Empirical Approach

This section gives a very brief description of the empirical approach adopted by Fack et al. (2019) to estimate student preferences under different assumptions. They consider a logit-type random utility model for the formal framework described in Section 2 where students are matched to schools through the (student-proposing) deferred acceptance algorithm. Besides the submitted ranked-ordered lists and the final matchings, the researcher observes school priorities and capacities.

The utility functions of the students are allowed to take any value on the real line, and student i's utility from being matched to school s is parameterised as:

$$\begin{aligned} u_{is} &= \sum_{\ell=2}^{m} \mathbf{1}_{\{s=\ell\}}(s,\ell)\,\beta_{\ell-1} + \text{nearest}_{is}\beta_m + \text{distance}_{is}\beta_{m+1} + \epsilon_{is} \\ &= Z_{is}'\beta + \epsilon_{is} \end{aligned}$$

Here, $\mathbf{1}_{\{s=\ell\}}(s,\ell), \ell = 2,3,\ldots,m$ are indicator variables for all but the first school in the dataset[6]. nearest_{is} indicates whether school s is the closest school to the student's residential address and distance_{is} captures the geographical walking distance between student i and school s. These covariates are collected in the $(m+1)\times 1$ vector Z_{is} and the product of this vector with the unknown model coefficients will henceforth be referred to as $V_{is} := Z_{is}'\beta$. Unobservable student heterogeneity ϵ_{is} is assumed to be drawn from a type 1 extreme value distribution. Note that attention is restricted to student preferences being independent of the placement of other students (i.e., no peer-effects) and that matching-specific statistics (e.g., cutoffs) do not enter the utility function.

The estimation relies on revealed student preferences in the data. What information that is revealed depends on the underlying assumption about if the weak truth-telling or the stability strategy is used. Suppose, for example, that there are four schools, called s_1, s_2, s_3 and s_4 with cutoffs 0.8, 0.9, 0.4 and 0.2, respectively. If student i with priority 0.5 at each school reports the ranked-ordered list $L_i = (s_1, s_3)$, it can be concluded that $s_1 \succ_i s_3 \succ_i s_j$ for $j \in \{2,4\}$ under the weak truth-telling assumption, but (ii) only that $s_3 \succ_i s_4$ under the stability assumption (since

[6] This implements the identifying restriction $\beta_1 = 0$ for the conditional logit models used to estimate student preferences.

school s_2 is non-feasible for student i).

3.1 Estimators

The type of information revealed under either weak truth-telling or stability assumptions motivates the corresponding maximum likelihood estimator of Fack et al. (2019) for the parameters in student-school utilities u_{is}. Under a weak truth-telling assumption, a student submitting a rank-order list of $L_i = (L_{i1}, \ldots, L_{ik_i})$ reveals $u_{L_{i1}} > \cdots > u_{L_{ik_i}}$. Furthermore, non-ranked schools are at the bottom of a student's utility profile, implying $u_{L_{ik_i}} > u_{is'}$. The likelihood of observing the entire rank-order list is modelled as the product of k_i marginal probabilities of observing school L_{ij} as $j = 1, 2, \ldots, k_i$-th ranked choice on the list. Each marginal probability is then further modelled as a conditional logistic regression whose choice alternatives are the observed choice for the current position on the rank-order list as well as all lower and non-ranked schools. The resulting log likelihood of such a chain of conditional logistic regressions is given by:

$$\ln L^{\mathrm{WTT}}(\beta, Z, |\phi^{\mathrm{WTT}}|) = \sum_{i \in I} \sum_{s \in \phi_i^{\mathrm{WTT}}} V_{is} - \sum_{i \in I} \sum_{s \in \phi_i^{\mathrm{WTT}}} \ln \left[\sum_{s' \not\succ_{\phi_i^{\mathrm{WTT}}} s} \exp\left(V_{is'}\right) \right].$$

The vector that maximizes this function will in the following be referred to as $\hat{\beta}^{\mathrm{WTT}}$.

Under a stability assumption, information about student preferences emerges from the observed assignment of students to schools rather than the submitted rank-order lists. In particular, admission of student i to school s implies that $u_{is} > u_{is'}$ for all other schools $s' \neq s$ that would admit student i given the realized student-school matching. Hence, the observed matching of students to schools can be seen as resulting from a conditional logistic regression with student-specific choice alternatives. The estimator resulting from maximum likelihood estimation of this model is henceforth denoted $\hat{\beta}^{\mathrm{ST}}$.

3.2 The Hausman Test

As demonstrated by Fack et al. (2019, p. 1506), the assumption of weak truth-telling nests the stability assumption for students that are matched with a school. Accordingly, testing weak truth-telling against stability boils down to testing whether the restrictions of weak truth-telling in excess of those imposed by stability hold. Fack et al. (2019) investigate this question indirectly via a Hausman test on the estimates of the coefficients β in student-school utilities u_{is}. Formally, the test statistic is given by:

$$T^H = \left(\hat{\beta}^{\text{ST}} - \hat{\beta}^{\text{WTT}}\right)' \left(\hat{\mathbf{V}}^{\text{ST}} - \hat{\mathbf{V}}^{\text{WTT}}\right)^{-1} \left(\hat{\beta}^{\text{ST}} - \hat{\beta}^{\text{WTT}}\right),$$

where $\hat{\mathbf{V}}^{\text{ST}}$ and $\hat{\mathbf{V}}^{\text{WTT}}$ are the estimated covariance matrices of $\hat{\beta}^{\text{ST}}$ and $\hat{\beta}^{\text{WTT}}$, respectively. Assuming that at least the stability assumption is satisfied and that the parametric assumptions of conditional logistic regressions are met, T^H has power against violations of weak truth-telling that render $\hat{\beta}^{\text{WTT}}$ inconsistent.

4 Data and Results

Schooling becomes compulsory in Sweden for children from the autumn term of the year they reach the age of six (grade 0), and compulsory school attendance ceases at the end of the spring term of their 10th school year, i.e., by the time they are 16 years old. A new legislation from 1992 (*Friskolereformen*) made it easier to start and operate voucher schools and, currently, around 16 percent of the children in grades 0 to 9 are enrolled in such schools. The municipality and voucher schools can, but need not, coordinate their admissions.[7]

This replication study is based on 75 school admissions in 15 different Swedish school districts, for grades 0, 4 and 7, between years 2019–2023 (different years for different districts). In all 75 data sets, we have data on all municipality schools (capacities, cutoffs, etc.) and how the students in the district rank them. In 37 of the data sets, we also have access to all such data for the voucher schools. In this case, the students submit one rank-ordered list containing both municipality and voucher schools. In 33 of the data sets, we don't have access to any information related to the voucher schools, and in 5 data set we have partial information about the voucher schools. See column three of Table 3 for details. In 71 of the 75 data sets, there are no restrictions on how many schools the students are allowed to rank. See column five of Table 3 for details,

[7]For more on the matching practices of the Swedish school system, see Andersson (2017).

where "x out of y" means that students are allowed to rank at most x of the y schools in the district. All schools that are included in a given data set are part of the same admission system and the admissions are always based on the (student proposing) deferred acceptance algorithm.

Table 1: Aggregated summary statistics for the 75 data sets.

	Total	Max	Mean	Min
Total number of students	78,445	4,007	1,046	131
Students with rank-ordered lists (ROL)	66,547	3,426	887	116
Number of schools	1,688	65	23	4
Number of seats (adjusted)	75,088	3,749	1,001	123
Number of placements	61,168	3,135	816	90
Number of first-choice placements	50,633	2,449	675	88

Table 1 contains some summary statistics of the data. Here, the "number of placements" refers to the number of students that have been assigned a school that is part of their ranked-ordered list (in reality, students are always placed at some school, possibly a non-ranked school). For each data set, the analysis is based only on the number of students that ranked at least one school in the school district ("Students with ranked-ordered lists") and not the "Total number of students." Students that did not submit ranked-ordered lists have been removed from the data set and initial capacities at each school have been reduced by the corresponding share. After that, a new placement has been made using the deferred acceptance algorithm, resulting in new (and modified) cutoffs. The analysis is then based on the reduced data sets and the assignment generated in them in order to emulate a centralized school choice system with only active students.

Note also that not all school districts use the same rules to determine their priorities. The rules can be based on a variety of variables, e.g., sibling priority, distance to alternative schools, right for some students to be placed in a school that is administratively connected to their current school, and so on. The Appendix contains all these details for all school districts. For our purposes, however, it suffices that the priorities are strict, possibly after a random tie-breaker has been applied (see Table **??**), independently of how they are created. Both priority scores and distance to schools have been normalized to belong to the closed interval $[0, 1]$.

4.1 Results

Estimation of student preferences under both weak truth-telling and stability assumptions as well as model testing have been conducted using the Matlab code in the replication package of Fack et al. (2019).[8] The P-values of a Hausman test for weak truth-telling versus stability assumptions in all 75 data sets are reported in Table 3, together with some additional characteristics of the data sets. We correct for multiple hypothesis testing using the sequential method of Holm (1979) and therefore only report adjusted P-values.

Test results with a targeted family-wise error rate of 5 percent (i.e., adjusted P-value threshold of 0.05) provide strong support for the empirical findings in Fack et al. (2019). More specifically, the null hypothesis of weak-truth telling is rejected in 66 of the investigated 75 data sets (88.0 percent). Table 4 summarises these findings and separates them in two dimensions (access to data from voucher schools and the size of the school districts) to investigate if they influence our findings, i.e., the rejections of the Hausman tests. As can be seen in the table, both access to data from voucher schools and larger student populations naturally increase the likelihood to reject the null-hypothesis.

To further investigate if students "skip the impossible," we go beyond a pure replication of the results in Fack et al. (2019) and estimate a logistic regression for ranking a school highest in the submitted ranked-ordered list. Formally, we specify:

$$\log\left(\frac{P(Y_{is}=1)}{1-P(Y_{is}=1)}\right) = \beta_0 + \beta_1 \text{infeasible}_{is} + \beta_2 \text{distance}_{is} + \sum_{\ell=1}^{m} \mathbf{1}_{\{s=\ell\}}(s,\ell)\,\beta_{\ell+2}.$$

Here, $Y_{is} \in \{0,1\}$ is a dummy indicating if school s is ranked as the top-choice by student i, $\text{infeasible}_{is} \in \{0,1\}$ is a dummy indicating if school s is ex-post infeasible and distance_{is} is the normalized walking distance between student i and school s. Lastly, $\beta_3, \beta_4, \ldots, \beta_{m+2}$ are school fixed effects. The specified logit model is estimated after merging the observations in all 75 data sets into a large stacked data set. Accordingly, school fixed effects apply to a school in a particular dataset so that schools in the same municipality are allowed to have different unobserved characteristics across admission years and for admissions to grade 0 and 7, respectively.[9]

[8]The replication package is available at `www.aeaweb.org/articles?id=10.1257/aer.20151422`.

[9]Analogous to previous model specifications, identification requires setting the fixed effect on the first school in the stacked dataset equal to zero.

Our formal test for "skipping the impossible" is a significance test on β_1, the coefficient on ex-post non-feasibility. If students did not take feasibility into account when submitting their ranked-ordered list, this test should fail to reject its null hypothesis since Y_{is} and infeasible$_{is}$ would be independent. As reported in Table 2, the null hypothesis of this test is rejected at a significance level of 0.5 percent (threshold chosen in accordance to Benjamin et al., 2018). Furthermore, the estimate of β_1 implies a difference in log odds ratios of -1.16 between feasible and non-feasible schools. Accordingly, the odds of ranking an infeasible school highest in the submitted ranked-ordered list are merely 21.1 percent of the odds of doing this for a feasible school.

Table 2: Estimation results for "skipping the impossible."

Coefficient	Estimate	Std. Error	P-value
Infeasible	-1.155	0.113	0.000^{***}
Distance	-17.402	0.605	0.000^{***}
Fixed effects	1,417 (school × dataset)		
N	2,061,193		

Note: Estimation conducted after stacking data in all 75 datasets. See Table 3 for a complete list. 7 fixed-effects (6,344 observations) were removed because of no first-choice rankings. Distance is normalised to 1.

Table 3: Data characteristics and the Hausman Test

Municipality (year)	Grade	Vouchers included?	Students with ROL	ROL restriction?	P-value
Huddinge (2023)	0	No	1,213	No	0.000*
Huddinge (2023)	7	No	574	No	0.000*
Järfälla (2020)	0	No	833	No	0.000*
Järfälla (2021)	0	Yes (all)	897	No	0.000*
Järfälla (2021)	7	Yes (all)	542	No	0.013*
Järfälla (2022)	0	No	894	No	0.000*
Järfälla (2022)	7	No	583	No	0.000*
Järfälla (2023)	7	No	661	No	0.100

Continued on next page

Table 3 – continued from previous page

Municipality (year)	Grade	Vouchers included?	Students with ROL	ROL restriction?	P-value
Karlstad (2021)	7	Yes (2 out of 4)	497	No	0.000*
Karlstad (2023)	7	Yes (all)	820	No	0.000*
Kävlinge (2021)	0	No	372	No	0.700
Kävlinge (2022)	0	No	441	No	0.000*
Kävlinge (2023)	0	No	391	Yes (3 out of 9)	0.001*
Linköping (2021)	0	Yes (all)	1,772	No	0.000*
Linköping (2021)	7	Yes (all)	1,590	No	0.000*
Linköping (2022)	0	Yes (all)	1,772	No	0.000*
Linköping (2022)	7	Yes (all)	1,767	No	0.000*
Linköping (2023)	7	Yes (all)	1,718	No	0.000*
Linköping (2023)	0	Yes (all)	1,589	No	0.000*
Malmö (2021)	0	No	3,275	Yes (7 out of 60)	0.000*
Malmö (2021)	7	No	1,239	Yes (7 out of 32)	0.000*
Malmö (2022)	0	Yes (3 out of 18)	3,426	No	0.000*
Malmö (2022)	7	Yes (2 out of 16)	1,398	No	0.000*
Nacka (2019)	0	Yes (all)	128	No	0.249
Nacka (2020)	0	Yes (all)	1,435	No	0.000*
Nacka (2020)	7	Yes (all)	734	No	0.000*
Nacka (2021)	0	Yes (all)	1,434	No	0.000*
Nacka (2021)	7	Yes (all)	698	No	0.000*
Nacka (2022)	0	Yes (all)	1,264	No	0.000*
Nacka (2022)	7	Yes (all)	412	No	0.000*
Nacka (2023)	0	Yes (all)	1,402	No	0.000*
Nacka (2023)	7	Yes (all)	602	No	0.000*
Norrköping (2020)	0	No	1,352	No	0.000*
Norrköping (2020)	7	No	947	No	0.000*
Norrköping (2021)	0	Yes (all)	1,513	No	0.000*

Continued on next page

Table 3 – continued from previous page

Municipality (year)	Grade	Vouchers included?	Students with ROL	ROL restriction?	P-value
Norrköping (2021)	7	No	959	No	0.000*
Norrköping (2022)	7	Yes (all)	907	No	0.000*
Norrköping (2023)	0	Yes (all)	1,490	No	0.000*
Norrköping (2023)	7	Yes (all)	1,126	No	0.000*
Sigtuna (2020)	7	No	336	No	0.565
Sigtuna (2021)	0	No	433	No	0.030*
Sigtuna (2021)	7	No	324	No	0.000*
Sigtuna (2022)	7	No	325	No	0.534
Sigtuna (2023)	0	No	546	No	0.005*
Sigtuna (2023)	7	No	424	No	0.000*
Trelleborg (2021)	0	No	417	No	0.076
Trelleborg (2021)	7	No	186	No	0.077
Trelleborg (2022)	0	No	513	No	0.000*
Trelleborg (2022)	7	No	116	No	0.045*
Trelleborg (2023)	7	No	138	No	0.229
Tyresö (2020)	0	Yes (all)	614	No	0.006*
Tyresö (2020)	7	Yes (all)	292	No	0.001*
Tyresö (2022)	0	Yes (all)	603	No	0.005*
Tyresö (2023)	0	Yes (all)	547	No	0.000*
Upplands-Bro (2020)	0	Yes (1 out of 3)	349	No	0.013*
Upplands-Bro (2021)	0	Yes (all)	330	Yes (at least 3)	0.007*
Upplands-Bro (2022)	0	Yes (all)	322	No	0.178
Upplands-Bro (2023)	0	Yes (all)	317	No	0.018*
Uppsala (2021)	0	Yes (all)	2,594	No	0.000*
Uppsala (2021)	7	Yes (all)	486	No	0.000*
Uppsala (2022)	0	Yes (all)	2,515	No	0.000*
Uppsala (2022)	7	Yes (all)	672	No	0.000*

Continued on next page

Table 3 – continued from previous page

Municipality (year)	Grade	Vouchers included?	Students with ROL	ROL restriction?	P-value
Uppsala (2023)	0	Yes (all)	2,482	No	0.000*
Uppsala (2023)	7	Yes (all)	721	No	0.000*
Växjö (2020)	7	No	734	No	0.000*
Växjö (2021)	7	Yes (2 out of 5)	797	No	0.000*
Växjö (2022)	7	Yes (all)	896	No	0.000*
Växjö (2023)	7	Yes (all)	939	No	0.000*
Ystad (2020)	7	No	271	No	0.000*
Ystad (2021)	0	No	276	No	0.016*
Ystad (2021)	7	No	279	No	0.000*
Ystad (2022)	0	No	283	No	0.001*
Ystad (2022)	7	No	279	No	0.000*
Ystad (2023)	0	No	251	No	0.013*
Ystad (2023)	7	No	273	No	0.000*

Table 4: Summary of results for the 75 data sets.

	Number of data sets	Data sets with *	Share with *
All data sets	75	66	88.0%
Data sets with no voucher data	33	26	78.8%
Data sets with voucher data	42	40	95.2%
Data sets with $\geq$ 500 students	47	46	97.9 %
Data sets with $<$ 500 students	28	20	71.4%

References

Abdulkadiroğlu, A. and Andersson, T. (2023). School choice. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume VI, chapter 3, pages 135–185. Elsevier, Amsterdam.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005a). The New York City high school match. *American Economic Review*, 95:364–367.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2005b). The Boston public school match. *American Economic Review, Papers and Proceedings*, 95:368–371.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93:729–747.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the New York City high school match. *American Economic Review*, 107:3635–3689.

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review*, 110:1502–1539.

Agarwal, N. and Somaini, P. (2018). Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86:391–444.

Andersson, T. (2017). Matching practices for elementary schools–Sweden. Matching in Practice, MiP Country Profile 24, `www.matching-in-practice.eu`.

Benjamin, D. J., Berger, J. O., Johannesson, M., and et al. (2018). Redefine statistical significance. *Nature Human Behavior*, 2:6–10.

Beuermann, D. W., Jackson, C. K., Navarro-Sola, L., and Pardo, F. (2022). What is a good school, and can parents tell? Evidence on the multidimensionality of school output. *The Review of Economic Studies*, 90:65–101.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What parents want: School preferences and school choice. *Economic Journal*, 125:1262–1289.

Che, Y.-K., Hahm, D. W., and He, Y. (2023). Leveraging uncertainties to infer preferences: Robust analysis of school choice. *Unpublished Working Paper*.

Chen, Y. and Sönmez, T. (2006). School choice: An experimental study. *Journal of Economic Theory*, 127:202–231.

Combe, J., Dur, U. M., Tercieux, O., Terrier, C., and Unver, M. U. (2022a). Market design for distributional objectives in (re)assignment: An application to improve the distribution of teachers in schools. *Unpublished Working Paper*.

Combe, J., Tercieux, O., and Terrier, C. (2022b). The design of teacher assignment: Theory and evidence. *The Review of Economic Studies*, 89:3154–3222.

de Haan, M., Gautier, P. A., Oosterbeek, H., and van der Klaauw, B. (2023). The performance of school assignment mechanisms in practice. *Journal of Political Economy*, 131:388–455.

Dur, U. M., Hammond, R. G., and Morrill, T. (2018). Identifying the harm of manipulable school choice mechanisms. *American Economic Journal: Economic Policy*, 10:187–213.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109:1486–1529.

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

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144:1921–1947.

Hahm, D. W. and Park, M. (2022). A dynamic framework of school choice: Effects of middle schools on high school choice. *Unpublished Working Paper*.

Hakimov, R. and Kübler, D. (2021). Experiments on centralized school choice and college admissions: A survey. *Experimental Economics*, 24:434–488.

Hastings, J., Kane, T. J., and Staiger, D. (2009). Heterogeneous preferences and the efficacy of public school choice. *Unpublished Working Paper.*

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from Two Experiments. *The Quarterly Journal of Economics*, 123:1373–1414.

He, Y. (2017). Gaming the Boston school choice mechanism in Beijing. *Unpublished Working Paper*.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Otero, S., Barahona, N., and Dobbin, C. (2021). Affirmative action in centralized college admission systems: Evidence from Brazil. *Unpublished Working Paper*.

Rees-Jones, A. and Shorrer, R. (2023). Behavioral economics in education market design: A forward-looking review. *Journal of Political Economy Microeconomics*, 1:447–613.

Romero-Medina, A. (1998). Implementation of stable solutions in a restricted matching market. *Review of Economic Design*, 3:137–147.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7:617–628.