// NO.22-046 | 10/2022

DISCUSSION PAPER

// JOSUE ORTEGA AND THILO KLEIN

Improving Efficiency and Equality in School Choice

Improving Efficiency and Equality in School Choice

Josué Ortega[a], Thilo Klein[b,c]

[a]*Queen's University Belfast, UK.*
[b]*ZEW – Leibniz Centre for European Economic Research, Mannheim, Germany.*
[c]*Pforzheim University, Germany.*

Abstract

How should students be assigned to schools? Two mechanisms have been suggested and implemented around the world: deferred acceptance (DA) and top trading cycles (TTC). These two mechanisms are widely considered excellent choices because they are strategy-proof, in addition to DA's no justified envy and TTC's Pareto optimality. We show theoretically and empirically that both mechanisms perform poorly with regard to two key desiderata such as efficiency and equality, even in large markets. In contrast, the rank-minimizing mechanism (RM) is significantly more efficient and egalitarian. It is also Pareto optimal for the students, unlike DA, and generates less justified envy than TTC.

Keywords: school choice, inequality, efficiency, justified envy.
JEL Codes: C78, D73.

Email address: `j.ortega@qub.ac.uk (corresponding author).`

We thank the anonymous referees of the MATCH-UP 2022 workshop (where a preliminary version of this paper was accepted) for their valuable suggestions. We also acknowledge helpful comments from Mustafa Afacan, David Delacrétaz, Aytek Erdil, Yoan Hermstrüwer, Jörgen Kratz, Aditya Kuvalekar, Vincent Meisner, Antonio Miralles, Hervé Moulin, Afshin Nikzad, Robert Parviainen, Juan Sebastián Pereyra, Erel Segal-Halevi, Olivier Tercieux, Peter Troyan, Bertan Turhan, Utku Ünver, Mark Wilson, Bumin Yenmez, Gabriel Ziegler, and audiences at the 2023 Econometric Society European Summer Meeting, Match-UP, the Conference on Economic Design, the Coalition Theory Network Workshop, the Belfast Easter Workshop on School Choice and seminars at ISER Essex and Boston College. We thank Sarah Fox and Taylor Knipe for proofreading this paper. We acknowledge funding from the ESRC and the Leibniz Association as part of project K125/2018: "Improving school admissions for diversity and better learning outcomes".

October 19, 2022

1. Introduction

School choice is a common way to assign students to schools based on the students' and schools' preferences. Students and schools rank their potential matches and submit this information to a centralized clearinghouse. Afterwards, an algorithm (also known as a mechanism) is applied to the submitted data and an allocation of students to schools is generated.

But which mechanism should we use to assign students to schools? Che and Tercieux (2018) convincingly argue that: "*the selection must be based on some measure of aggregate welfare of participants. For instance, if one Pareto efficient mechanism yields a significantly higher utilitarian welfare level or a much more equal payoff distribution than others, that would constitute an important rationale for favoring such a mechanism*".

We show that the mechanism that minimizes the sum of ranks for the students (henceforth RM, Featherstone, 2020) outperforms two of the most popular mechanisms used in school choice with respect to the two desiderata named above, i.e. utilitarian welfare and equality. RM is superior to Gale's and Shapley's deferred acceptance (DA) mechanism and to Gale's top trading cycles (TTC) mechanism in that: i) RM assigns the average student to a school they prefer more (i.e. it is more efficient), and ii) RM assigns the worst-off student to a school that they prefer *much* more (i.e. it is more egalitarian).

In particular, if there are n students and n schools with one seat each, and preferences for both sides are drawn uniformly at random, TTC and DA asymptotically assign the average student to approximately their $\log(n)$ most preferred school, whereas RM assigns them to a school better than their second choice. If we focus on the worst placement, rather than the average, the difference is even bigger: RM assigns them to their $\log_2(n)$ most preferred school, whereas DA assigns them to their $\log^2(n)$ and TTC to a school in the bottom half of their rank list (see Fig. 1 for the rank distribution). In addition, because all Pareto optimal and strategy proof mechanisms are equivalent (TTC and serial random dictatorship included), our results show that requiring strategy-proofness in school choice mechanisms can significantly hurt efficiency and equality.

Furthermore, RM is Pareto optimal for the students, unlike DA, and generates justified envy for fewer students than TTC, which is surprising because RM does not use schools' priorities but TTC does. We prove these properties for random markets where preferences are drawn independently and uniformly at random (see Table 1), and document them by analyzing real data from the student assignment system in Budapest (see Fig. 2).

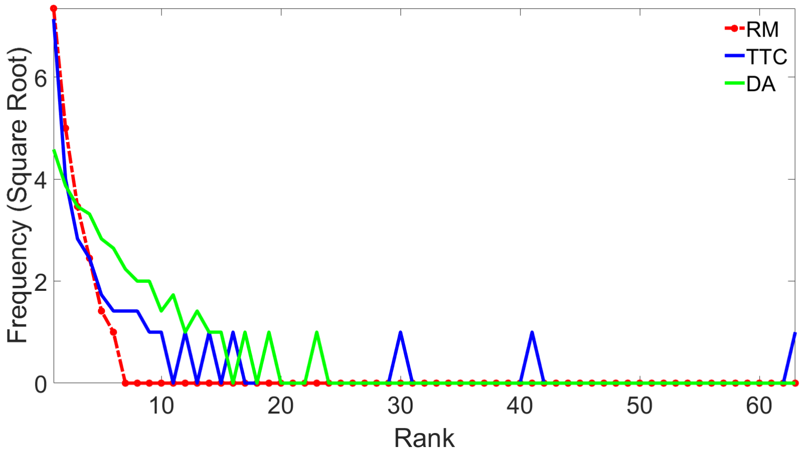


Figure 1: Rounded average rank distribution in 1,000 random markets with $n = 100$. Preferences are drawn independently and uniformly at random. The x-axis is truncated at the highest value with positive density. See Appendix B for details.

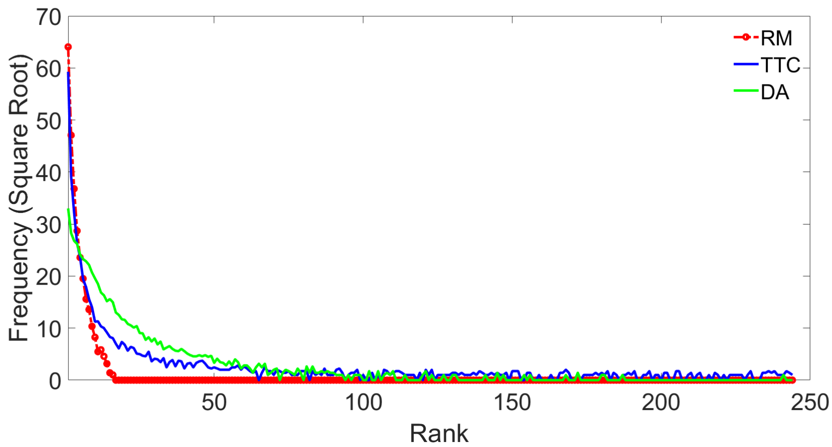


Figure 2: Rank distribution generated for 10,131 students in the secondary school admissions in Budapest. The maximum rank is 244. See Section 5 for details.

Table 1: Theoretical properties of school choice mechanisms in large random markets.

	RM	TTC	DA
Average rank	≤ 2	$\log(n)$	$\log(n)$
Maximum rank	$\log_2(n)$	$> 0.5\,n$	$\log^2(n)$
Students w. justified envy	$0.33\,n$	$0.39\,n$	0
Pareto optimal	Yes	Yes	No
Strategy-proof	No	Yes	Yes

2. Related Literature

The question of which mechanism should be used to assign students to schools has been frequently asked. The answer to this question in the market design literature is that all frequently used mechanisms generate equivalent rank distributions. This equivalence has been established theoretically for a wide class of mechanisms (Che and Tercieux, 2018; Pycia, 2019) and empirically using real-life data (Abdulkadiroğlu et al., 2009; Pathak and Sönmez, 2013; Che and Tercieux, 2018; Abdulkadiroğlu et al., 2020). Our paper challenges the literature consensus by showing that Pareto optimal mechanisms are not equivalent, as can be observed in Figs. 1 and 2.[1] Three reasons explain the discrepancy between our results and those in the literature, namely i) different model specifications, ii) we consider non-strategy proof mechanisms, like RM, and iii) RM has not been used in empirical studies. We explain these differences in detail below.

The closest paper to ours is Che and Tercieux (2018). Using a random market approach, they show that the normalized payoff distribution generated by any Pareto optimal mechanism is asymptotically equivalent. Furthermore, they compare the rank distribution generated by DA and TTC (but not RM) using data from the New York City school choice program. The main lesson from their paper is that all Pareto optimal mechanisms are equivalent in large markets, and therefore there is no reason to prefer any Pareto optimal mechanism over another. Our paper shows that the payoff equivalence between Pareto optimal mechanisms breaks down once i) ranks are used instead of normalized payoffs, and ii) students are allowed to rank all available schools, rather than just a few.

Pycia (2019) obtains a similar equivalence result to that of Che and Tercieux: he shows that any anonymous statistics, such as rank distribution, generated by Pareto efficient and strategy-proof mechanisms are equivalent, even in finite markets (note that RM is not strategy-proof).[2] This implies that all of our results for TTC's poor performance with regards to efficiency and equality also apply to the random serial dictatorship mechanism (RSD), which "*has a long history and is used in a wide variety of practical allocation problems, including school choice, worker assignment, course allocation, and the allocation of public housing*" (Pycia and Troyan, 2021).

To show that RM is more efficient than DA and TTC, we connect the school

[1] A two-sample Kolmogorov-Smirnov test rejects the null-hypothesis that any two distributions in Figures 1 or 2 are the same at the 1% significance level.

[2] Pyicia's result builds on a previous, more general equivalence result for large markets by Liu and Pycia (2016).

choice problem to that of assigning one of n jobs to each of n workers so to minimize costs.[3] Worker i incurs a cost c_{ij} when completing job j. The matrix C contains all such costs. When each row of C is an independent random permutation of $\{1, \ldots, n\}$, this problem is equivalent to that of finding the rank-minimizing allocation of students to schools, ignoring schools' priorities. Each entry c_{ij} denotes the rank (cost) of school (job) j for student (worker) i. To show that the RM is more efficient than TTC and DA, we invoke a result in Parviainen (2004) which shows that the cost-minimizing allocation has an average cost smaller than 2, and compare it with the well-known average rank in TTC and DM, which is around $\log(n)$. Obtaining the maximum rank lower bound and the fraction of students with justified envy is easy using the limit distribution of ranks in RM, which is also provided by Parviainen.

The result of average rank being bounded in RM was recently independently discovered by Nikzad (2022), who provides a bound of 7.75 (ours is 2). His proof uses random graph arguments and is different (and significantly more involved) than ours. Sethuraman (2022) shows that Nikzad's bound can be improved to 2 using the cost assignment problem with costs distributed in $(0, 1)$ (Aldous, 2001), without using Parviainen's result. These papers do not study the maximum rank and justified envy in RM, TTC and DA, and do not analyse the performance of these three mechanisms using real-life data in which preferences are correlated.

The RM mechanism was first studied in economics by Featherstone (2020). He documents that RM has been used in practice to assign teachers to schools in the US, and shows that any selection of the RM mechanism cannot be strategy-proof. Nonetheless, he shows that truth-telling is a best response in RM when students have little information about other students' preferences and do not truncate their preference list. He shows that a rank efficient allocation must be ordinally efficient (and thus ex-post efficient), but the converse is not necessarily true. He also shows that an inefficient assignment can converge to the RM outcome by performing local swaps. Troyan (2022) has recently shown that RM is non-obviously manipulable, meaning that although potential manipulations exist, they cannot be recognized by cognitively limited agents. Therefore, RM has better icentives proprties than the well-known Boston mechanism, which is obviously manipulable.

The fact that DA is inefficient is well-known: Kesten (2010) shows that, in a

[3] A large literature in mathematics, uncited in economics, has studied this problem. See Olin (1992) and Krokhmal and Pardalos (2009) for a summary of it.

worst-case scenario, it may assign each student to her worst or second-worst school. We show that DA is also inefficient in an average-case scenario. The inefficiency of TTC is less known, partially because the matching literature often focuses on the weaker efficiency notion of Pareto optimality. Nonetheless, Manea (2009) has shown that the number of preference profiles for which RSD is ordinally efficient (and thus rank efficient) vanishes when the number of agents grows. Our result complement his by showing that RSD (and TTC) not only rarely produces a rank efficient allocation, but also the size of its inefficiency does not vanish in large markets. To our knowledge, the inequality of both mechanisms has remained largely unstudied in the economics literature.

TTC minimizes justified envy among all Pareto optimal and strategy-proof mechanisms (Abdulkadiroğlu et al., 2020). Neither DA nor RM are in this class of mechanisms. We find theoretically that fewer students experience justified envy in RM than in TTC. In practice RM and TTC generate roughly the same amount of justified envy.

3. Model

We study a standard one-to-one school choice market (Abdulkadiroğlu and Sönmez, 2003), which consists of:

1. A set of students $T = \{1, \ldots, n\}$,
2. A set of schools $S = \{s_1, \ldots, s_n\}$, with each school having space for one student only,
3. Strict students' preferences over schools $\succ := (\succ_1, \ldots, \succ_n)$, and
4. Strict schools' priorities over students $\triangleright = (\triangleright_{s_1}, \ldots, \triangleright_{s_n})$.

An allocation x is a perfect matching between T and S. We will denote by x_t the school to which student t is assigned, and by x_s the student that school s is assigned to. Student i experiences justified envy in allocation x if there exists a school s such that $s \succ_t x_t$ and $t \triangleright_s x_s$.

The function $\mathrm{rk}_t(x_t)$ returns an integer between 1 and n corresponding to the ranking of x_t in the preference list of student t, i.e. the most desirable option gets a ranking of 1, whereas the least desirable one gets a ranking of n. A mechanism is a map from school choice markets to (a probability distribution over) allocations. An allocation x Pareto dominates a different allocation y if, for every student t, $\text{rk}_t(x_t) \leq \text{rk}_t(y_t)$ and for some student j, $\text{rk}_j(x_j) < \text{rk}_j(y_j)$. An allocation is Pareto optimal if it is not Pareto dominated. A Pareto optimal mechanism returns a Pareto optimal allocation in every school choice problem.

We use x^{RM} to denote one of the (possibly many) allocations that minimizes the sum of ranks for students, which we henceforth call rank efficient or rank minimizing. X^{RM} denotes the set of all rank efficient allocations. The rank-minimizing mechanism (henceforth RM) is one that returns a rank efficient allocation for every matching market.[4]

Two other mechanisms are of interest. The first is top trading cycles (TTC), in which the following two steps are repeated until all agents have been assigned an object:

1. Construct a graph with one vertex per student or school. Each student (resp. school) points to their top-ranked school (resp. student) among the remaining ones. At least one cycle must exist and no two cycles overlap. Select the cycles in this graph.
2. Permanently assign each student in a cycle to the school they point to. Remove all students and schools involved in a cycle.

The second mechanism of interest is student-proposing deferred acceptance (DA). It works as follows:

1. All unmatched students apply to their most preferred school that has not rejected them. Each school that has received a proposal puts the one sent by the highest priority student in a waiting list and permanently rejects all other received applications (if any).
2. Repeat step 1 until all schools have received at least one application. Assign each student to the school which has them on a waiting list.

We use x^{TTC} and x^{DA} to denote the allocation obtained by the TTC and DA mechanisms, respectively. Schools' priorities are used to compute TTC and DA, but are irrelevant in RM.

[4]Rank efficiency is a stronger efficiency notion than ordinal efficiency and Pareto optimality (Featherstone, 2020). We simply write efficiency to refer to rank efficiency.

4. Results

Our theoretical results relate to the properties of the expected allocation generated by RM, TTC and DA when students' preferences and schools' priorities are drawn independently and uniformly at random. This assumption is commonly used to analyze matching markets.[5] We study the asymptotic behavior of: i) expected average rank (efficiency), ii) expected maximum rank (inequality), and iii) expected number of students with justified envy generated by RM, TTC and DA in the next subsections.[6]

Efficiency. We first study the expected average rank generated by RM, TTC and DA in random markets. To do so, we define $\overline{x} := \frac{1}{n}\sum_{i=1}^{n} \mathrm{rk}_i(x_i)$, which denotes the average rank of the school to which students are assigned in allocation x.

Proposition 1 shows that the expected average ranking in RM is smaller (i.e. better) than that in TTC and DA. It follows directly from a result by Parviainen (2004) that has not yet been cited in the economics literature. In contrast, the results for DA and TTC are well-known and we simply restate them for completeness.

Proposition 1. *The expected average rank in RM, TTC and DA is:*

$$\lim_{n\to\infty} \mathbb{E}[\overline{x}^{RM}] \leq 2 \tag{1}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}[\overline{x}^{TTC}]}{\log n} = 1 \tag{2}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}[\overline{x}^{DA}]}{\log n} = 1 \tag{3}$$

Proof. Statement 1 is proven by Parviainen (2004, Theorem 2.1, p. 105).[7] Statement 2 is proven by Knuth (1996, equation 4, p. 439).[8] Statement 3 is proven by Pittel (1989, Theorem 2, p. 538).

[5]See Che and Tercieux (2018) and references therein.

[6]Expected ranks are commonly used in matching markets to measure welfare (e.g. Ashlagi et al., 2017).

[7]Parviainen (2004) also provides a lower bound, and thus the expected average rank in RM is such that $\pi^2/6 \leq \lim_{n\to\infty} \mathbb{E}[\overline{x}^{\mathrm{RM}}] \leq 2$.

[8]Knuth shows that $\mathbb{E}[\sum_{i=1}^{n} \mathrm{rk}_i(x_i^{\mathrm{TTC}})] = (n+1)H_n - n$, where H_n is the n-th harmonic

□

Proposition 1 shows that the rank inefficiency of DA and TTC does not vanish as the market grows large because, even if the average rank obtained by DA and TTC grows slowly with the size of the market, the average rank obtained by RM is constant and does not grow with n.

Inequality. We measure inequality as the rank of the object obtained by the worst-off agent in the market, i.e. the maximum rank in the rank distribution. This measure follows John Rawls' idea that the welfare of a society is that of its worst-off member.[9] To do so, we define $\underline{x} \coloneqq \max_i \mathrm{rk}_i(x_i)$, which denotes the rank of the object obtained by the worst-off agent in allocation x.

Proposition 2 shows that RM generates a significantly more egalitarian allocation than DA and TTC. In particular, TTC generates an allocation so unequal that the worst-off student is assigned to a highly undesirable school in the lower half of their preference list. Such rank is much higher than the corresponding value for RM ($\log_2(n)$) and DA ($\log^2(n)$).

Proposition 2. *The expected maximum rank of RM, TTC and DA is:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[\underline{x}^{RM}]}{\log_2(n)} = 1 \tag{4}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}[\underline{x}^{TTC}]}{n} > 0.5 \tag{5}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}[\underline{x}^{DA}]}{\log^2(n)} = 1 \tag{6}$$

Proof. Statement 5 was proven by Knuth (1996, p. 440). Statement 6 was proven by Pittel (1992), theorem 6.1, p. 382 and note before references, p. 400.

number and, therefore, $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\sum \mathrm{rk}(x_i^{\mathrm{TTC}})] = \log(n)$. See also the note after the acknowledgements in Frieze and Pittel (1995), p. 807.

[9] Alternatively, one could define inequality as the difference in ranks between the worst- and best-off agent. Because the rank of the object obtained by the best-off agent is 1 in any Pareto optimal allocation (Abdulkadiroğlu and Sönmez, 1998), both measures are equivalent.

To prove statement 4 we use the asymptotic rank distribution in RM. The probability that a student is assigned to their i-th choice is asymptotically equal to $\frac{1}{2^i}$ (Theorem 1.3 in Parviainen (2004)), so that a student is assigned to a school with rank 1 with probability 1/2, to a school with rank 2 with probability 1/4, and so on. This probability distribution is very similar to the number of consecutive heads in n independent coin tosses, in which 0 heads obtains with probability 1/2, 1 heads with probability 1/4 and so on (the distribution of ranks in RM is shifted by +1). Finding the longest run of heads is a known problem, in which the longest run is equal to $\frac{\log(n/2)}{\log(2)}$ (Schilling, 2012). Therefore, the maximum rank in RM is equal to $\frac{\log(n)/2}{\log(2)} + 1 = \log_2(n)$.[10] □

Although we only provide a lower bound for the maximum rank in TTC (of $0.5\,n$), simulations suggest that the maximum rank in TTC converges to $0.63\,n$.

Justified Envy. We use e^{RM}, e^{TTC} and e^{DA} to denote the fraction of students who experience justified envy in the allocation obtained in RM, TTC and DA, respectively. Proposition 3 shows that RM generates fewer cases of expected envy than TTC, which is interesting since TTC is envy minimal in the class of strategy-proof and Pareto optimal mechanisms (Abdulkadiroğlu et al., 2020).

Proposition 3. *The expected fraction of students with justified envy in RM, TTC and DA is:*

$$\lim_{n\to\infty} \mathbb{E}[e^{RM}] = 0.33 \tag{7}$$

$$\lim_{n\to\infty} \mathbb{E}[e^{TTC}] = 0.3863 \tag{8}$$

$$\lim_{n\to\infty} \mathbb{E}[e^{DA}] = 0 \tag{9}$$

Proof. Statement 9 is well-known, as DA does not generate justified envy (Gale and Shapley, 1962).

[10] Frieze and Sorkin (2007, Theorem 2, p. 1436) proves that the maximum cost in the cost assignment problem when costs are uniformly distributed in $[0, n]$ is $\Theta(\log(n))$.

For the remainder of the proof we use the fact that the number of students with justified envy in TTC (e^{TTC}) and RSD (e^{RSD}) is asymptotically equivalent (Che and Tercieux, 2017). Since schools' priorities are irrelevant in both RM and RSD, a student who is assigned to their i-th most preferred school does *not* experience justified envy with probability $\frac{1}{2^{i-1}}$. To see this, notice that students placed into their 1st choice trivially do not experience justified envy with probability 1; students placed into their second best choice do not experience justified envy if the student who is accepted at his most preferred school has a higher priority than them, which occurs with probability 1/2; for students who are assigned to their third choice, they do not experience justified envy if their first and second most preferred school rank their assigned student above them, i.e. with probability 1/4, and so on.

Thus, to obtain the total fraction of students who do not experience justified envy in RM and RSD (TTC), we just need to multiply i) the probability that a student matched to their i-th most preferred school does not experience justified envy, times ii) the fraction of students who are assigned to such a choice in RSD and RM. The fraction of students assigned to their i-th choice in RM asymptotically equals $\frac{1}{2^i}$ (Theorem 1.3 in Parviainen (2004)), whereas in RSD the probability that the k-th dictator is assigned to his j-th most preferred school is given by $p_{k,j} = \frac{1}{n!}\binom{k-1}{j-1}(j-1)!(n-j)!(n+1-k)$ (Knuth, 1996).[11] Putting these expressions together, and after some algebra for the RSD case detailed in Appendix A, we obtain:

$$e^{\mathrm{RM}} = 1 - \sum_{i=1}^{n} \frac{1}{2^i} \times \frac{1}{2^{i-1}} = 1 - \sum_{i=1}^{n} \frac{1}{2^{2i-1}} \to 0.33 \tag{10}$$

$$e^{\mathrm{RSD}} = 1 - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{i} \frac{1}{n!} \binom{k-1}{j-1} (j-1)!(n-j)!(n+1-k) \frac{1}{2^{i-1}} \to 0.3863 \tag{11}$$

which finalizes the proof, since $\lim_{n\to\infty} \mathbb{E}[e^{\mathrm{TTC}}] = \lim_{n\to\infty} \mathbb{E}[e^{\mathrm{RSD}}]$. $\square$

[11] For example, if $k = 1$, then $p_{1,1} = 1$ and $p_{1,j} = 0$ for any $j > 1$. Similarly, when $j = 1$, then $p_{k,1} = \frac{n+1-k}{n}$. Note that $\binom{n}{0} = 1$ and $\binom{n}{m} = 0$ for any $m > n$.

5. Data

One critique that can be made to our random market results is that they assume that students' preferences are independent, whereas students' preferences tend to be correlated, and such correlation may improve the performance of DA and TTC with regards to efficiency and equality. We show that this is not the case by using real-life data from secondary school admissions in Hungary in 2015. In summary, we find that TTC and DA perform even worse than when we assumed independent uniform preferences.

Our data contains the preferences and priorities of 10,131 students and 244 schools in Budapest. Because students only rank a few schools and schools only rank students who apply to them, we apply RM, DA and TTC to i) the actual reported preferences and priorities (Hungary assigns students to schools using DA (Biró, 2008), so using the reported preferences as real preferences is a sensible strategy), and ii) the estimated, complete preferences and priorities.[12] Figures 2 and 3 present the distribution of the ranks realized after applying RM, DA and TTC to the estimated and reported preferences, respectively. Table 2 presents summary statistics.

Table 2: Rank descriptive statistics for Budapest.

Preferences	Estimated Preferences			Reported Preferences		
Variable \ Mechanism	RM	TTC	DA	RM	TTC	DA
Mean	2.7	8.9	12.3	1.5	1.9	2.1
Maximum	16	244	241	6	14	13
Variance	4.7	607.4	220.1	0.6	1.7	1.9
Share of students w. justified envy	0.58	0.64	0	0.46	0.44	0
Unassigned students	–	–	–	2,555	2,508	2,704

For reported preferences, the mean is computed dividing by the number of assigned students.

The lessons we learn from computing the rank distributions in Budapest are similar to those we learned from looking at random markets. Table 2

[12]The complete preferences are estimated using the original students' reported preferences assuming that i) students do not use dominated strategies, and ii) the realized assignment is stable. For the complete estimation procedure see Aue et al. (2020). In both cases, we balanced the demand and supply for seats by adjusting the schools capacities (this is because there are substantially more school seats than students, repeating the exercise with the unbalanced market does not change the comparison between mechanisms). When a student only ranks k schools, we use $k+1$ as the rank of being unassigned. RM chooses the rank minimizing assignment randomly among all rank efficient allocations.

shows that RM performs better than TTC and the currently used DA with regards to efficiency and equality with reported and estimated preferences. When full, estimated rank lists are used, the average student substantially improves their placement (average rank in RM is 2.7, compared to 8.9 in TTC and 12.3 in DA). RM still generates a better average rank when we use stated preferences, but the difference with the average rank generated by TTC and DA is smaller (this is because the average student only ranks 4.1 schools on average). RM assigns the average student to a school in their 16 percentile of their preference lists, whereas the corresponding percentile for DA and TTC are 35 and 29, respectively.

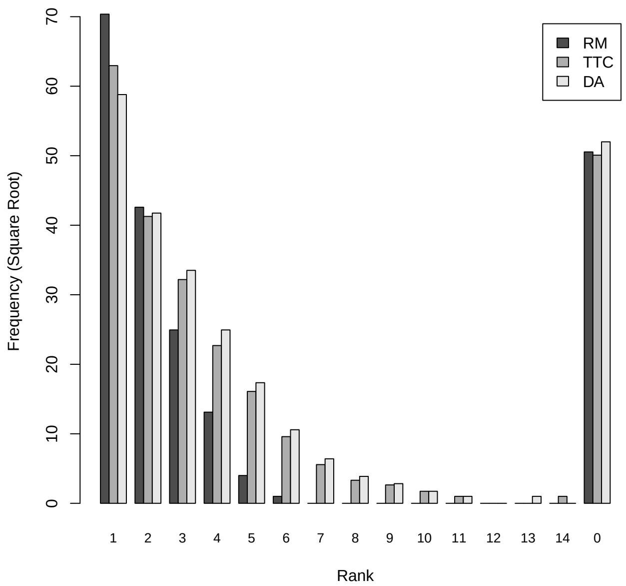


Figure 3: Rank distribution generated for 10,131 students in the secondary school admissions in Budapest using reported preferences. The last bar (0) denotes unassigned students.

With regards to inequality, RM performs much better than DA and TTC with complete preferences, assigning the worst-off student to the 16th best choice rather than to their 241th and 244th, respectively (out of 244). It also assigns less than 2% of the student population to their 10th ranked school or worse, whereas TTC and DA assign 16% and 41% of the student population to such school, respectively. RM also generates a significantly more egalitarian allocation with reported (incomplete) preferences, assigning the worst-off student to the 6th best choice rather than their 13th or 14th best. With estimated and reported preferences, we find that DA and TTC are incomparable in terms of equality, since TTC assigns more students to a really undesirable school, but also assigns more students to a top 3 school. The number of students unassigned in RM, TTC and DA is roughly equivalent when reported (incomplete) preferences are used.

Our rank distributions are similar to those documented in other studies. Che and Tercieux (2018) and Abdulkadiroğlu et al. (2020) also document that TTC assigns more students to their first choice than DA. Both studies also find that DA and TTC generate a similar number of unassigned students.

Our empirical analysis has used the preferences that students submit in DA (arguably their true ones) to generate the TTC and RM allocations. One potential concern is that students would submit different preferences when allocations are determined by RM, which is not strategy-proof. To tackle this concern, we computed the rank distribution generated by RM when a fraction of the students who have incentives to misrepresent their preferences do so. We have found that the rank distribution and number of students with justified envy remain largely unchanged, and thus is safe to conclude that the rank distribution generated by RM is more efficient and equal than the ones generated by DA and TTC, even when a fraction of agents misrepresent their preferences.[13]

Moreover, RM is non obviously manipulable, and thus cannot be manipulated by cognitively limited agents (Troyan, 2022). Furthermore, the potential gains from manipulation are small (the average student can only improve by less than one rank in their preference list with iid preferences, and by less than 2 ranks in the data). Finally, there is evidence that truthful behavior can be more common in non strategy-proof mechanisms than in strategy-proof ones (Cerrone et al., 2022).[14] Thus, whether students would actually misrepresent their preferences in RM more than they do in DA or TTC is an interesting open question.[15]

[13]See the Appendix for detailed summary statistics.

[14]Cerrone et al. (2022) find that almost twice as many people (70% versus 40%) behave truthfully in the efficiency adjusted deferred acceptance (EADA) mechanism versus standard DA, even though DA is strategy-proof and EADA is not.

[15]Substantial strategic behavior in strategy-proof mechanisms such as DA and TTC has been documented in the lab (Chen and Sönmez, 2006; Rees-Jones and Skowronek, 2018; Hakimov and Kübler, 2021; Guillen and Veszteg, 2021; Cerrone et al., 2022) and in real life (Hassidim et al., 2017; Rees-Jones, 2018; Shorrer and Sóvágó, 2018).

6. Conclusion

Our paper highlights the efficiency loss and inequality generated by the celebrated deferred acceptance and top trading cycles mechanisms.

October 19, 2022

References

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2009): "Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match," *American Economic Review*, 99, 1954–78.

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

Abdulkadiroğlu, A., Y.-K. Che, P. A. Pathak, A. E. Roth, and O. Tercieux (2020): "Efficiency, justified envy, and incentives in priority-based matching," *American Economic Review: Insights*, 2, 425–42.

Abdulkadiroğlu, A. and T. Sönmez (1998): "Random serial dictatorship and the core from random endowments in house allocation problems," *Econometrica*, 66, 689–701.

Aldous, D. J. (2001): "The ζ (2) limit in the random assignment problem," *Random Structures & Algorithms*, 18, 381–418.

Allen, R., S. Burgess, and L. McKenna (2013): "The short-run impact of using lotteries for school admissions: early results from Brighton and Hove's reforms," *Transactions of the Institute of British Geographers*, 38, 149–166.

Ashlagi, I., Y. Kanoria, and J. Leshno (2017): "Unbalanced random matching markets: The stark effect of competition," *Journal of Political Economy*, 125, 69–98.

Aue, R., T. Klein, and J. Ortega (2020): "What happens when separate and unequal school districts merge?" *Centre for European Economic Research Discussion Paper.*

Biró, P. (2008): "Student admissions in Hungary as Gale and Shapley envisaged," *University of Glasgow Technical Report TR-2008-291.*

Bogomolnaia, A. and H. Moulin (2001): "A new solution to the random assignment problem," *Journal of Economic Theory*, 100, 295 – 328.

Brown, M., C. Donnelly, P. Shevlin, C. Skerritt, G. McNamara, and J. O'Hara (2021): "The rise and fall and rise of academic selection: The case of Northern Ireland," *Irish Studies in International Affairs*, 32, 477–498.

Cerrone, C., Y. Hermstrüwer, and O. Kesten (2022): "School choice with consent: an experiment," *MPI Collective Goods Discussion Paper*.

Che, Y.-K. and O. Tercieux (2017): "Top trading cycles in prioritized matching: An irrelevance of priorities in large markets," *Unpublished Manuscript, Columbia University*.

——— (2018): "Payoff equivalence of efficient mechanisms in large matching markets," *Theoretical Economics*, 13, 239–271.

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

Coldron, J., E. Tanner, S. Finch, L. Shipton, C. Wolstenholme, B. Willis, S. Demack, and B. Stiell (2008): "Secondary school admissions," Tech. rep., Department for Children, Schools and Families.

Featherstone, C. (2020): "Rank efficiency: Modeling a common policymaker objective," *Unpublished paper, University of Pennsylvania*.

Frieze, A. and B. G. Pittel (1995): "Probabilistic analysis of an algorithm in the theory of markets in indivisible goods," *The Annals of Applied Probability*, 768–808.

Frieze, A. and G. B. Sorkin (2007): "The probabilistic relationship between the assignment and asymmetric traveling salesman problems," *SIAM Journal on Computing*, 36, 1435–1452.

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

Guillen, P. and R. F. Veszteg (2021): "Strategy-proofness in experimental matching markets," *Experimental Economics*, 24, 650–668.

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

Hassidim, A., D. Marciano, A. Romm, and R. I. Shorrer (2017): "The mechanism is truthful, why aren't you?" *American Economic Review*, 107, 220–24.

Kesten, O. (2010): "School choice with consent," *The Quarterly Journal of Economics*, 125, 1297–1348.

Knuth, D. (1996): "An exact analysis of stable allocation," *Journal of Algorithms*, 20, 431 – 442.

Krokhmal, P. A. and P. M. Pardalos (2009): "Random assignment problems," *European Journal of Operational Research*, 194, 1–17.

Liu, Q. and M. Pycia (2016): "Ordinal efficiency, fairness, and incentives in large markets," *SSRN preprint*.

Manea, M. (2009): "Asymptotic ordinal inefficiency of random serial dictatorship," *Theoretical Economics*, 4, 165–197.

Moulin, H. (2003): *Fair division and collective welfare*, Cambridge, MA, USA: MIT Press.

Nikzad, A. (2022): "Rank-optimal assignments in uniform markets," *Theoretical Economics*, 17, 25–55.

Noden, P., A. West, and A. Hind (2014): "Banding and ballots: secondary school admisssions in England: admissions in 2012/13 and the impact of growth of academies," Tech. rep., The Sutton Trust.

Olin, B. (1992): *Asymptotic properties of random assignment problems*, PhD Thesis, Stockholm Royal Institute of Technology.

Oosterbeek, H., N. Ruijs, and I. De Wolf (2020): "Using admission lotteries to estimate heterogeneous effects of elite schools," *Unpublished paper, SSRN.*

Parviainen, R. (2004): "Random assignment with integer costs," *Combinatorics, Probability and Computing*, 13, 103–113.

Pathak, P. A. and T. Sönmez (2013): "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation," *American Economic Review*, 103, 80–106.

Pittel, B. (1989): "The average number of stable matchings," *SIAM Journal on Discrete Mathematics*, 2, 530–549.

——— (1992): "On likely solutions of a stable marriage problem," *The Annals of Applied Probability*, 358–401.

Pycia, M. (2019): "Evaluating with statistics: Which outcome measures differentiate among matching mechanisms?" *Unpublished paper, University of Zurich.*

Pycia, M. and P. Troyan (2021): "A theory of simplicity in games and mechanism design," *Unpublished paper, University of Zurich.*

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

Rees-Jones, A. and S. Skowronek (2018): "An experimental investigation of preference misrepresentation in the residency match," *Proceedings of the National Academy of Sciences*, 115, 11471–11476.

Schilling, M. F. (2012): "The surprising predictability of long runs," *Mathematics Magazine*, 85, 141–149.

Sethuraman, J. (2022): "A note on the average rank of rank-optimal assignments," *unpublished.*

Shapley, L. and H. Scarf (1974): "On cores and indivisibility," *Journal of Mathematical Economics*, 1, 23 – 37.

Shorrer, R. I. and S. Sóvágó (2018): "Obvious mistakes in a strategically simple college admissions environment: Causes and consequences," *Available at SSRN 2993538.*

Troyan, P. (2022): "Non-Obvious Manipulability of the Rank-Minimizing Mechanism," *arXiv preprint arXiv:2206.11359.*

Appendix A - Proof of Proposition 3

In RSD the probability that the k-th dictator is assigned to his j-th most preferred school is given by

$$p_{k,j} = \frac{1}{n!}\binom{k-1}{j-1}(j-1)!(n-j)!(n+1-k) \tag{12}$$

In equation (12), if $k = 1$, then $p_{1,1} = 1$ (the probability that the first dictator gets his first school is one) and $p_{1,j} = 0$ for any $j > 1$ (the probability that the first dictator gets a school worse than his top one is zero). Similarly, when $j = 1$, then $p_{k,1} = \frac{n+1-k}{n}$ (this is the probability that the k-th dictator gets his top school, or equivalently, the probability that the $k-1$ dictators before him are assigned to the school that dictator k ranks as first). Note that $\binom{n}{0} = 1$ and $\binom{n}{m} = 0$ for any $m > n$, so that the probability that dictator k is assigned to a school with a rank higher than k is zero.

Equation (12) can be rewritten as:

$$\begin{aligned} p_{k,j} &= (n+1-k)\frac{(k-1)!(j-1)!(n-j)!}{n!(j-1)!(k-j)!} && (13)\\ &= (n+1-k)\frac{(k-1)!(j-1)!(n-j)!}{n!(j-1)!(k-j)!}\frac{k}{k}\frac{j}{j} && (14)\\ &= \frac{(n+1-k)}{k}\frac{\binom{k}{j}}{\binom{n}{j}} && (15)\end{aligned}$$

Since RSD is independent of schools' priorities, a student placed in their j-th most preferred school does *not* experience envy with probability $\frac{1}{2^{j-1}}$. Therefore, the total number of students without justified envy in RSD (NE^{RSD}) equals

$$\begin{aligned} NE^{\mathrm{RSD}} &= \sum_{k=1}^{n}\sum_{j=k}^{n}\frac{(n+1-k)}{k}\frac{\binom{k}{j}}{\binom{n}{j}}\frac{1}{2^{j-1}} && (16)\\ &= \sum_{j=1}^{n}\frac{1}{\binom{n}{j}2^{j-1}}\sum_{k=j}^{n}\frac{(n+1-k)}{k}\binom{k}{j} && (17)\\ &= \sum_{j=1}^{n}\frac{1}{\binom{n}{j}2^{j-1}}\left[(n+1)\underbrace{\sum_{k=j}^{n}\frac{1}{k}\binom{k}{j}}_{A} - \sum_{k=j}^{n}\binom{k}{j}\right] && (18)\end{aligned}$$

Where

$$A = \sum_{k=j}^{n} \frac{1}{k}\binom{k}{j} = \sum_{k=j}^{n} \frac{1}{k}\frac{k!}{j!(k-j)!} = \sum_{k=j}^{n} \frac{(k-1)!}{j\,(j-1)!(k-j)!} \tag{19}$$

$$= \frac{1}{j}\sum_{k=j}^{n}\binom{k-1}{j-1} = \frac{1}{j}\sum_{k=j-1}^{n-1}\binom{k}{j-1} \tag{20}$$

Plugging this in our expression for NE^{RSD}, we have

$$NE^{\mathrm{RSD}} = \sum_{j=1}^{n} \frac{1}{\binom{n}{j}2^{j-1}}\left[\frac{n+1}{j}\sum_{k=j-1}^{n-1}\binom{k}{j-1} - \sum_{k=j}^{n}\binom{k}{j}\right] \tag{21}$$

Using the Hockey-stick identity $\sum_{k=j}^{n}\binom{k}{j} = \binom{n+1}{j+1}$, we obtain

$$NE^{\mathrm{RSD}} = \sum_{j=1}^{n} \frac{1}{\binom{n}{j}2^{j-1}}\left[\frac{n+1}{j}\binom{n}{j} - \binom{n+1}{j+1}\right] \tag{22}$$

$$= \sum_{j=1}^{n} \frac{1}{\binom{n}{j}2^{j-1}}\left[\frac{n+1}{j}\binom{n}{j} - \frac{(n+1)\,n!}{(j+1)\,j!(n-j)!}\right] \tag{23}$$

$$= \sum_{j=1}^{n} \frac{1}{\binom{n}{j}2^{j-1}}\left[(n+1)\left(\frac{1}{j} - \frac{1}{j+1}\right)\binom{n}{j}\right] \tag{24}$$

$$= (n+1)\sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j-1}\left(\frac{1}{j} - \frac{1}{j+1}\right) \tag{25}$$

We divide both sides by $n+1$ to obtain

$$\frac{NE^{\text{RSD}}}{n+1} = \sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j-1}\frac{1}{j} - \sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j-1}\frac{1}{j+1} \tag{26}$$

$$= 2\sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j}\frac{1}{j} - 4\sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j+1}\frac{1}{j+1} \tag{27}$$

$$= 2\sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j}\frac{1}{j} - 4\sum_{j=2}^{n+1}\left(\frac{1}{2}\right)^{j}\frac{1}{j} \tag{28}$$

$$= -2\sum_{j=2}^{n}\left(\frac{1}{2}\right)^{j}\frac{1}{j} + 1 - \frac{4}{(n+1)}\left(\frac{1}{2}\right)^{n+1} \tag{29}$$

$$= -2\sum_{j=2}^{n}\left(\frac{1}{2}\right)^{j}\frac{1}{j} + 1 - \frac{4}{(n+1)}\left(\frac{1}{2}\right)^{n+1} - 1 + 1 \tag{30}$$

$$= -2\sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j}\frac{1}{j} + 2 - \frac{4}{(n+1)}\left(\frac{1}{2}\right)^{n+1} \tag{31}$$

$$= 2 - \underbrace{\frac{1}{(n+1)}\left(\frac{1}{2}\right)^{n-1}}_{B} \underbrace{-2\sum_{j=1}^{n}\left(\frac{1}{2}\right)^{j}\frac{1}{j}}_{C} \tag{32}$$

As n goes to infinity, B goes to 0 and C is the Taylor expansion for $-\ln(1/2)$.

$$\lim_{n\to\infty}\frac{NE^{\text{RSD}}}{n} = \lim_{n\to\infty}\frac{NE^{\text{RSD}}}{n+1} = \lim_{n\to\infty} 2 + 2\ln(\frac{1}{2}) = 0.6137 \tag{33}$$

Thus, the fraction of students who experience justified envy in RSD tends to $e^{\text{RSD}} = 1 - 0.6137 = 0.3863$, which is what we wanted to prove. We thank Peter Košinár for suggesting this approximation.

Appendix B - Simulations

In simulated markets (see Tables 3 and 4), we clearly see that RM dominates TTC and DA in efficiency (average rank) and inequality (maximum rank). Given the large ranks that realize in TTC, it is unsurprising that the variance of the rank distribution is large too. The variance of RM is much smaller, which shows that the ranks are heavily concentrated among the first four top choices. Table 3 also allows us to assess the accuracy of the random market results presented in section 4. For TTC, the mean rank is surprisingly close to the theoretical prediction (± 1 of $\log(n)$). In RM, the upper bound provided of 2 for the mean is quite tight, and the approximation $\log_2(n)$ for the max rank is also remarkably accurate.

Table 3: Rank descriptive statistics. Average over 1,000 simulations.

Variable	$n = 100$			$n = 500$		
	RM	TTC	DA	RM	TTC	DA
Mean	1.8	4.3	5.0	1.8	5.8	6.7
Max	6	64	23.2	8	315	42.6
Variance	1.3	73.3	18.5	1.34	421.0	37.6

The severity of the inequality generated by TTC is fully exposed in Table 4. TTC not only makes someone really worse off, assigning them a really bad object ($0.63\,n$), but it assigns an object in the bottom 90% (not top 10%) of their preferences to over 1.5% of the agents. In contrast, RM does not assign such a poor option to any agent. RM also assigns more agents to their top choice than TTC.

Table 4: Percentage of agents who receive an object with rank higher (worse) than m. Average over 1,000 simulations.

	$n = 100$			$n = 500$		
m	RM	TTC	DA	RM	TTC	DA
1	46	50	96	46.6	49.8	93.2
2	21	33	72	21.0	33.2	79.6
$\log(n)$	3	20	40	0.6	14.2	39.8
$0.1\,n$	0	8	12	0	1.8	0
$0.25\,n$	0	3	0	0	0.6	0
$0.5\,n$	0	1	0	0	0.2	0

Appendix C - Strategic applicants

To understand the impact of manipulating students on the rank distribution generated by RM, we compute the RM allocation when a fraction (20%, 40%, 60% and 80%) of students who have incentives to manipulate do so. We consider two possible manipulations.

Drop-Assigned: In the reported preference data from Budapest, 5,183 students are not assigned their first preference in the RM. A first possible manipulation we consider is for them to move their assigned school to the end of their preference list in the hope of increasing admissions chances at a more preferred school. That is, a student who ranks $s_1 \succ \cdots \succ s_i \succ \cdots \succ s_n$ and gets s_i will perform a manipulation of the form $s_1 \succ \cdots \succ s_n \succ s_i$. This is equivalent to not ranking the school to which the applicant would have been assigned.

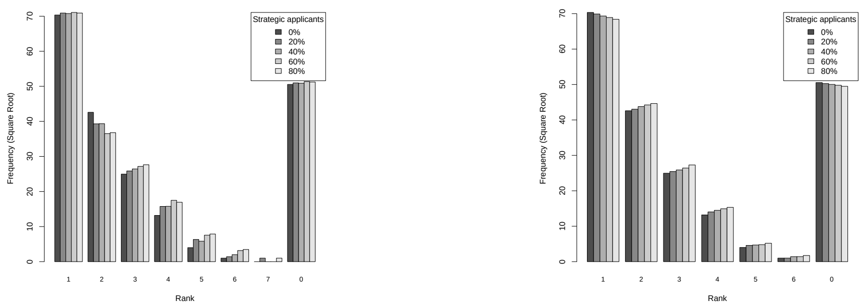
Drop-First: Another 3,369 students are neither assigned their first nor their second preference in the RM. A second possible manipulation we consider is for them to move their first preference to the end of their preference list in the hope of increasing admissions chances at their second preference. That is, a student who ranks $s_1 \succ s_2 \succ s_3 \succ \cdots \succ s_n$ and gets s_3 or worse will perform a manipulation of the form $s_2 \succ s_3 \succ \cdots \succ s_n \succ s_1$. This manipulation has been observed in real life applications (Abdulkadiroğlu and Sönmez, 1998).

Table 5 shows that the summary statistics in RM remain largely unchanged in the presence of strategic applicants. The statistics for the RM with a share of 0% strategic applicants are equivalent to the RM results with reported preferences in Table 2. With an increasing share of strategic applicants, the average and maximum rank change marginally, remaining well below the corresponding ranks for DA and TTC. Justified envy remains about the same. The number of unassigned students slightly decreases for the first manipulation, and slightly increases for the second one.

Figure 4 presents the rank distributions. Overall, the results show that the rank distributions remain largely unchanged, even for large shares of strategic applicants.

Table 5: Rank descriptive statistics for RM with strategic applicants.

Reported Preferences Variable \ RM mechanism	Share (number) of strategic applicants 0% (0)	 20% (1,036)	 40% (2,073)	 60% (3,109)	 80% (4,146)
Panel A: Drop-Assigned					
Mean	1.48	1.50	1.51	1.53	1.54
Maximum	6	7	6	6	7
Variance	0.58	0.70	0.70	0.81	0.82
Share of students w. justified envy	0.46	0.45	0.45	0.46	0.46
Unassigned students	2,555	2,599	2,590	2,636	2,625
Panel B: Drop-First					
Mean	1.48	1.50	1.52	1.54	1.56
Maximum	6	6	6	6	6
Variance	0.58	0.61	0.63	0.65	0.67
Share of students w. justified envy	0.46	0.46	0.46	0.47	0.47
Unassigned students	2,555	2,526	2,503	2,479	2,449



(a) *Drop-Assigned.*

(b) *Drop-First.*

Figure 4: Rank distribution generated by RM for 10,131 students in the secondary school admissions in Budapest using reported preferences by fraction of strategic applicants. The last bar (0) denotes unassigned students.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de