Personalized Recommendations Without Inducing Congestion: Mitigating Disparities in the NYC High School Match

ERICA CHIANG, Cornell University, USA
KENNY PENG, Cornell University, USA
REBECCA LICHTENSTEIN, New York City Public Schools, USA
BRIELLE MCDANIEL, New York City Public Schools, USA
KRISTEN O'NEIL, New York City Public Schools, USA
DEJAUNIQUE THOMAS, New York City Public Schools, USA
LIANNA WRIGHT, New York City Public Schools, USA
JON KLEINBERG, Cornell University, USA
EVA TARDOS, Cornell University, USA
NIKHIL GARG, Cornell University, USA

Algorithmic recommendations can help participants navigate large matching markets. For example, recommendations for school and college choices may reduce information frictions and disparities in access to high-performing programs. At scale, however, recommenders in capacity-constrained settings can be self-defeating: if they steer too many users toward the same items, then even users who were originally predicted to have a high chance of matching to an item may not, due to increased competition. In this paper, we formalize this phenomenon of *recommendation-induced congestion*; motivated by the NYC high school match, we show that naive recommendations can cause sharp decreases in program acceptance rates, most affecting applicants with the fewest nearby options. Next, we propose and theoretically analyze a congestion-aware, bilevel optimize-and-simulate approach to allocate recommendations and improve match outcomes *safely, in equilibrium*. Finally, we deploy this approach in the 2025-26 admissions cycle of the NYC high school match, aiming to reduce disparities by highlighting personalized lists of nearby, high-performing programs where an applicant has a high predicted offer likelihood. In a randomized controlled trial, we find that 16.4% of treatment applicants ranked a recommended program, versus 10.5% of control applicants who ranked a program they would have been recommended (57% relative increase; p=0.011); 5.6% of treatment applicants matched to such a program, versus 3.3% of control applicants (71% relative increase; p=0.071); further, no treatment applicant was rejected from a recommended program. Our findings suggest that recommenders should be analyzed and designed as market-shaping interventions.

October 4, 2026

1 INTRODUCTION

People navigate large matching markets to find educational opportunities, jobs, and partners. At scale, these markets can increase access to opportunity, but it may be difficult to actually identify suitable matches. In turn, many platforms use *recommender systems* to guide participants. While these recommenders are often designed to optimize individual-level outcomes, they have *systemic* effects—one participant's recommendations impose externalities on others, either their potential matches or competitors. In our specific context, the New York City (NYC) high school match, eighth-grade students apply to high school by submitting a ranked list of programs, and a centralized algorithm matches each student to one program [Abdulkadiroğlu et al., 2005, Gale and Shapley, 1962]. While the system was designed to expand access, forming preferences over programs is a complex, challenging process [Marian, 2023]. Students must choose from over 900 high school programs, which vary widely in academic focus and performance, transportation requirements, and extracurricular activities.

Forming informed preferences thus requires substantial time, information, and support. Given heterogeneous resources to navigate these informational burdens, the application complexities may perpetuate disparities [Broyard, 2023, Cohodes et al., 2025, Corcoran et al., 2018, Corradini, 2024, Corradini and Idoux, 2025b]. For example, Peng et al. [2026] find that Black, Hispanic, and low-income students in NYC disproportionately *undermatch*: they could have matched to programs that measured higher on certain metrics (program selectivity, academic performance, impact, and other measures of quality), while being at least as close to home as their true match, had they applied; the largest gaps were among the students with the most competitive applications.

Such disparities have motivated several informational interventions in NYC [Cohodes et al., 2025, Corcoran et al., 2018]. One prominent approach used in these interventions is to provide applicants at target middle schools with *non-personalized* lists of nearby, high-performing programs. Because every student within a given middle school receives the same list, such an intervention must either (a) be restricted to programs that most students would be admitted to (e.g., historically undersubscribed programs), or (b) risk steering many students to programs where they are ultimately rejected. The former limits the set of high schools that can be included, since many high-performing programs are competitive and require high grades and favorable lottery numbers. The latter can induce disappointment and may reduce trust in the intervention in future years.[1]

Recent advances in personalized predictions of admissions outcomes enable moving beyond these approaches, toward personalized guidance regarding where a student is likely to be admitted. Beginning in the 2024-25 cycle, in work by Corradini and Idoux [2025a], NYC Public Schools (NYCPS) has provided all high school applicants with personalized, program-specific offer likelihood estimates during the application process. These likelihoods are bucketed into {*High*, *Medium*, *Low*}, with applicants told that *High* corresponds to at least a 99% chance of admission and *Low* to a less than 1% chance [NYC Public Schools, 2026a]. Similar tools have recently been integrated into other school matching platforms [Arteaga et al., 2022b, Larroucau et al., 2024], including program-specific recommendations that depend on such predictions. More generally, predictions of downstream match outcomes are also increasingly incorporated into recommender systems in settings such as labor and dating platforms.

Recommendation-induced congestion. The above discussion overlooks an important concern: recommender systems are not passive information tools. Match outcomes are *endogenous* to all applicants' behavior, and an intervention's goal is to change behavior. Thus, individual-level interventions may be self-defeating: if many applicants are steered toward the same program, many—even those with *High* estimated offer likelihood—may be rejected due to new competition that was not accounted for in the original likelihood estimation. We refer to this phenomenon as *recommendation-induced congestion.* For example, Peng et al. [2026] identify individually-optimal application changes that would improve certain metrics of each student's matched program if adopted in isolation; however, they find in simulations that when many students adopt their "optimal changes" simultaneously, many students end up not matching at all. In this work, we further show in simulations that recommendation-induced congestion itself can be disparate: students with less access to high-performing programs are more likely to be affected by congestion at the programs they *are* eligible for, due to geographic correlation.

This work. We formalize the challenge of and propose an approach to allocating recommendations without inducing congestion. We apply our approach and leverage offer likelihood estimates to

[1] Corcoran et al. [2018] and Cohodes et al. [2025] navigate this tradeoff by providing all students a list of about 25 programs at varying selectivity levels with guidance on implementing a "reach, match, safety" approach [Ali and Shorrer, 2025]. Cohodes et al. [2025] also provide interactive search and recommendation tools using preferences such as for extracurriculars.

deploy a *personalized informational intervention*. In partnership with NYCPS, we sent applicants personalized lists of high-performing, nearby programs where they have *High* predicted offer likelihoods—focusing on applicants from middle schools that historically did not send many students to such schools. In addition to the overarching intervention goal of addressing disparities in applications to high-performing programs, a key design constraint in our deployment was to avoid recommendation-induced congestion. Our recommendations needed to preserve offer likelihood estimates even if many students adopted them; this was a non-negotiable constraint for our NYCPS partners, in order to prevent misleading applicants and to retain trust in the offer likelihood tool, future interventions, and the match process broadly. Our contributions are as follows.

Characterization of recommendation-induced congestion. Through empirically-calibrated simulations, we show that naively recommending each program to all eligible students can induce congestion at scale, failing to preserve offer likelihoods. The effects are heterogeneous, most affecting applicants with less *ex ante* access to high-performing programs, reinforcing disparities.

Method for safe recommendations in equilibrium. To mitigate such effects, we propose a bi-level optimization method for allocating recommendations in capacity-constrained settings, safely. We manage congestion by capping the number of nudges sent to each program, using historical data to determine how high nudge capacities can be while meeting congestion constraints. Then, at deployment time, we optimize nudge allocations given the nudge capacities. Theoretically, we show that under natural assumptions, our approach approximately maximizes expected desirable matches, while preserving acceptance rates. Empirically, our method achieves target acceptance rates without reducing overall intervention effectiveness compared to the naive approach.

Deployment and RCT of personalized recommendations. We deployed our method in an informational intervention for the 2025-26 NYC high school admissions cycle, providing applicants from target middle schools with *personalized* recommendations to nearby, high-performing programs they are likely to be admitted to even if recommendations are adopted at scale. We evaluate our intervention through a pre-registered randomized controlled trial (RCT), finding substantial take-up of our recommendations among treatment applicants: 16.4% of the 444 treatment applicants apply to a program they were nudged to, while 10.5% of the 822 control applicants apply to a program they would have been recommended—a 57% increase (p=0.011). We find suggestive evidence of an increase in *matches* to nudged programs: 5.6% of treatment applicants match to a nudged program, as opposed to 3.3% for control applicants, a 71% relative increase (p=0.071). Furthermore, no treatment applicants were rejected from a program they were nudged to,[2] and only one control applicant was rejected from a program they *would have* been nudged to.

We further apply our framework to study how big an effect such demand-side interventions can have. While our discussion in this paper is motivated by school placement due to our deployment, these considerations also arise in other capacity-constrained settings, such as labor and dating platforms: individual-level interventions in any of these settings should consider system-wide effects for safe, effective outcomes.

2 RELATED WORK

This work is motivated by substantial documentation of information frictions in school choice [Ainsworth et al., 2023, Bleemer and Zafar, 2018, Hastings et al., 2015, 2016, Jensen, 2010, Jochim et al., 2014, Walters, 2018, Wiswall and Zafar, 2015], the resulting suboptimal application behavior [Arteaga et al., 2022b, Larroucau et al., 2024, Marian, 2023, Tomkins et al., 2023], and the existence of *disparities* in behavior and outcomes [Broyard, 2023, Condliffe et al., 2015, Corradini, 2024, Corradini and Idoux, 2023, 2025b, Dynarski and Scott-Clayton, 2006, Garg, 2025, Hoxby and Avery,

[2]The dropoff from 16.4% to 5.6% is fully explained by matches to programs listed higher than the nudged program.

2012, Hoxby and Turner, 2015, Peng et al., 2026, Sattin-Bajaj et al., 2018, Singer, 2025]. More broadly, a large literature considers the design of school assignment systems and empirically evaluates their performance, including a focus on disparities [Allman et al., 2022, Angrist et al., 2024a, Awadelkarim et al., 2023, Dur et al., 2018, Immorlica et al., 2020, Jeloudar et al., 2025].

Informational interventions for school choice. Many informational interventions have aimed to increase awareness (and reduce disparities) in school assignment, both in NYC [Cohodes et al., 2025, Corcoran et al., 2018] and elsewhere [Agte et al., 2024, Ainsworth et al., 2020, Allende et al., 2019, Arteaga et al., 2022a,b, Campos, 2024, Elacqua et al., 2025, Hastings et al., 2007, Hastings and Weinstein, 2008, Larroucau et al., 2024, Neilson, 2026, Valant and Weixler, 2022, Ye, 2024]. They have provided students with information on school performance, value added, admissions chances, and/or outcomes. While some interventions deploy personalized search and recommendation tools, to our knowledge none of them have directly controlled recommendation-induced congestion. Most relevant to ours is the work of Corradini and Idoux [2025a], as discussed in the introduction; in NYC, their offer likelihood tool provides warnings to applicants in the online portal if they do not apply to any program where they have *High* likelihood. This is intended to prevent students from failing to match at all, as opposed to preventing undermatching. Arteaga et al. [2022a] provide program-specific recommendations that depend on such predictions. Larroucau et al. [2024] provide personalized *major* recommendations, but not *program* recommendations, and further include warnings if an applicant is highly likely to be admitted to their first-choice option (suggesting that they may consider listing a more competitive program first). Further, some of these systems use offer likelihoods or other characteristics for their *search* features.

A benefit of using personalized offer likelihoods is that we can include more competitive programs: in our deployment, performance cutoffs defining high-performing schools were higher than those used in prior NYC interventions, and over half of the recommended programs were oversubscribed the previous year (had more applicants than seats) and thus would have been excluded in a non-personalized intervention limited to programs where most would be admitted.

Other past work has analyzed the prediction of admissions outcomes in similar settings [Dong et al., 2026, Pathak and Shi, 2021]. In particular, Dong et al. [2026] show that machine learning-based predictions can become unstable when the applicant pool changes. Among other things, this implies the need to account for equilibrium effects when shifting application behavior in capacity-constrained allocation systems. Our work designs recommendations that explicitly account for these equilibrium effects when guiding applicants' choices.

Recommendation in equilibrium. While classical recommendation and decision support tools have historically been designed for uncapacitated environments, there is growing interest in accounting for scarcity, congestion, and fairness in recommendation settings. Recent work has developed methods for explicitly incorporating downstream matching and fairness considerations into recommendation design by balancing item exposure [Chen et al., 2024, 2025, Christakopoulou et al., 2017, Greenwood et al., 2024, Mashayekhi et al., 2023, Shi, 2025, Su et al., 2022]; used assortment optimization to dynamically show a restricted set of items to each user, while achieving final outcomes that maximize matches or revenue and are aligned with user preferences [Ashlagi et al., 2022, Housni et al., 2024, Rios and Torrico, 2023, Torrico et al., 2020]; studied mechanisms for reducing congestion by changing incentives [He and Magnac, 2022]; and modeled congestion in capacity-constrained settings to characterize design tradeoffs [Arnosti et al., 2021, Besbes et al., 2025, Bied et al., 2023, 2021, Escobar and Montes, 2025, Kanoria and Saban, 2021]. One setting with similar considerations of balancing allocations in a stage *before* a matching is the interview stage for medical residencies, with proposed approaches to reduce congestion [Allman et al., 2025, Ashlagi et al., 2020, 2025, Wapnir et al., 2021].

While the above papers are largely theoretical, several recommender deployments have been congestion aware or estimated congestion ex post. Rios et al. [2023] propose dynamic exposure policies and improve match rates in two dating markets, Manshadi et al. [2025] develop and deploy a model for rankings in a large volunteer matching platform, Huang et al. [2022] improve match efficiency in a dating market by controlling disclosure of demand information, Behaghel et al. [2024] deploy congestion-aware recommendations with French public employment services to maximize potential employment, and Chen et al. [2023] develop an equilibrium-aware recommendation approach and apply it to a large online dating service. Other recommender deployments report congestion effects: Neilson [2026] documents congestion from information and recommendation tools in Chile's school choice system; Altmann et al. [2026] provide online job search advice at scale and find that employment and earnings gains diminish when more people receive treatment; and Bobba et al. [2023] provide applicants with feedback about their own academic performance to drive skill-based sorting into high school tracks, using the experimental results to model scale-up and finding that congestion can have negative downstream effects, including on equity.

Our work builds on this literature by developing, deploying, and evaluating a method for recommendation in equilibrium under hard capacity and offer likelihood constraints.

Recommendation-induced congestion is also closely related to concerns of algorithmic monoculture in hiring and other matching markets, in which employers using the same hiring algorithms may induce congestion on who they interview and make offers to [Bommasani et al., 2026, 2022, Creel and Hellman, 2022, Kim et al., 2025, Kleinberg and Raghavan, 2021, Peng and Garg, 2024a,b]. Related work also studies the design of information displays in settings where agents respond strategically to shared information in the presence of congestion [Kayongo et al., 2021, Zhang et al., 2024]. Relatedly, recommendations based on offer likelihoods changing the offer likelihoods themselves is an example of performative prediction [Perdomo et al., 2020]. In Appendix A, we further discuss the relationship to work on experimentation in equilibrium.

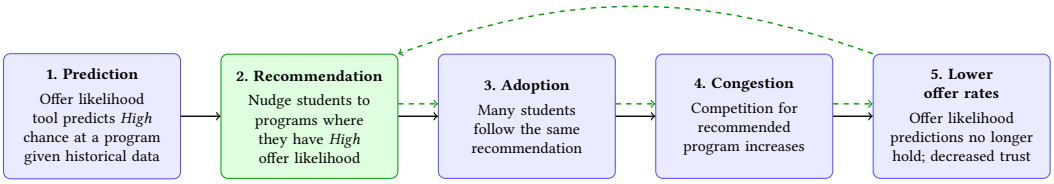
3 SAFE RECOMMENDATION IN CAPACITY-CONSTRAINED SETTINGS

The overarching goal of an intervention is to shift application behavior to increase matches among a specified set of "desirable matches." We consider interventions that achieve these shifts by nudging students to programs they would not otherwise have applied to. However, because programs have limited capacity, changing demand can also change acceptance rates. In Section 3.1, we characterize the consequences of failing to account for these shifts. In Section 3.2, we formalize our goal of safely allocating recommendations ("nudges") to achieve improved match outcomes without inducing congestion. Finally, in Section 3.3, we develop a solution for this concern—a heuristic bilevel optimization approach to compute such an allocation—and theoretically analyze its performance. Figure 1a overviews the problem and technical approach.

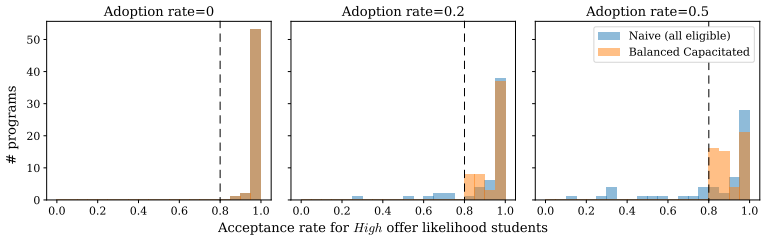
3.1 Motivation: consequences of recommendation-induced congestion

To illustrate recommendation-induced congestion, we simulate[3] the potential consequences of a naive approach to recommendation using data from the 2024-25 NYC high school admissions cycle. (Detailed descriptions of the data and setting appear in Section 4; here, we discuss results to motivate our method.) Recall that in our deployed intervention, our desirable matches are between students from target middle schools and nearby, high-performing high school programs where they

[3]In our setting, we can accurately *simulate* match outcomes given rankings. Then, using historical data on rankings and assumptions about how applicants respond to nudges, we can evaluate match outcomes for any nudge allocation.



(a) Our method simulates outcomes (green arrows) to generate recommendations that mitigate congestion.



(b) Simulated acceptance rates at 3 different simulated adoption rates.

Fig. 1. **Recommendation-induced congestion.** (a) shows a conceptual diagram of recommendation-induced congestion and how our method mitigates congestion. (b) shows acceptance rates at nudge-eligible programs, at various rates of simulated nudge adoption. Under **Naive** nudge allocation (blue), simulated acceptance rates among *High* offer likelihood students drop substantially with adoption. We compare these to our **Balanced Capacitated** nudge allocation approach designed to keep these rates $\geq 80\%$ (orange).

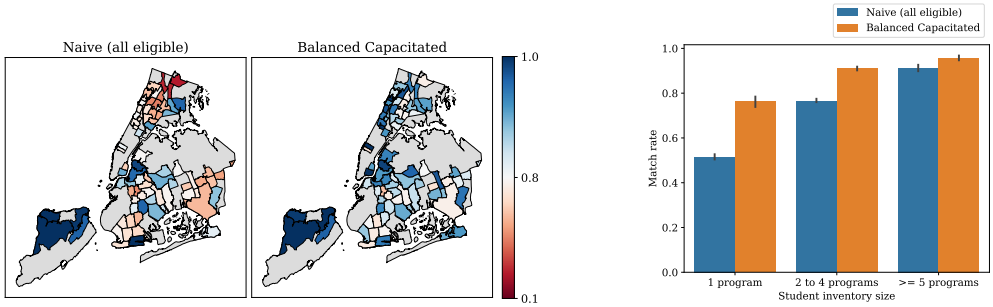
have *High* offer likelihood. The naive intervention approach nudges each student to *all* programs that meet these nudge-eligible criteria (i.e., all pairs forming a desirable match).[4]

Figure 1b shows histograms of simulated acceptance rates, among students with *High* pre-nudge offer likelihoods across programs, under different assumed rates of students adopting nudges. In blue is the naive approach of nudging all students to all eligible programs (we defer discussion of orange to Section 5). Adoption of nudges causes steep *drop-offs in program acceptance rates*, even among students who ex ante were told they have *High* (at least 99%) likelihood of being admitted.

Furthermore, congestion can induce *disparate effects* when applicants have unequal access to high-performing programs. In NYC, there are large geographic and income-based disparities in the number of high-performing programs accessible to applicants (illustrated in Figure 9a). Under a naive recommendation approach, congestion is disproportionately high at programs in areas where students tend to have less access to other options. Figure 2 illustrates this effect in simulation: applicants with less *ex ante* access to these programs have the lowest match rates to nudged programs, even when nudged.

Thus, naive nudge allocation would fail to preserve offer likelihoods, with disparate effects. This would undermine both trust in the offer likelihood tool, which is designed to help students more effectively navigate program options, and would also diminish the effectiveness of the intervention at scale. Our method is designed to address both of these effects.

[4]Appendix C.1 further evaluates the non-personalized approach used in prior interventions (every student at each middle school receives the same list of all nearby, high-performing programs), which induces even more congestion.



(a) Simulated match rates by zip code. Gray zip codes have ≤ 2 adopting students.

(b) Simulated match rates by inventory size (access to high-performing programs).

Fig. 2. **Match rates to nudged programs at 20% adoption.** In simulation, the **Naive** approach of nudging all eligible students to all eligible high school programs leads to *steep, disparate* drop-offs in the match rates of nudged students to their nudged programs, despite the fact that they originally had *High* offer likelihoods.

3.2 Model and Problem Definition

Suppose we have some exogenously-specified set of "desirable matches" Z, where each binary Z_{ij} indicates whether or not it would be desirable for student i to match to program j. The *intervention goal* is to increase the number of matches (i, j) with $Z_{ij} = 1$. In our setting, a student-program match is desirable ($Z_{ij} = 1$) only if it meets certain "nudge-eligibility" criteria: an applicant i attends a target middle school, has *High* offer likelihood at a high-performing program j, and lives near j. (Section 4 contains a detailed description.) The *recommendation task* is to choose an allocation of nudges X, where each binary X_{ij} indicates whether student i is nudged to program j. We only nudge individuals toward desirable matches ($X_{ij} \leq Z_{ij}$). Given a nudge allocation X, students adopt nudges and apply to programs according to a function $f(X)$. The resulting applications induce a match outcome $\mu_X = M(f(X))$, with $\mu_X(i)$ denoting applicant i's match.

Definition 1 (Congestion constraint). We require nudges to preserve offer likelihoods. For the realized match process, let $\ell_j(\mu) = \min\{\ell_j^{\text{high}}(\mu), \ell_j^{\text{nudged}}(\mu)\}$, where $\ell_j^{\text{high}}(\mu), \ell_j^{\text{nudged}}(\mu)$ are the acceptance rates at program j among all *High* likelihood applicants and the subset nudged to j, respectively.[5] Then, we require that $\ell_j(\mu_X) \geq L$ for all $j \in \mathcal{P}$, where $\mathcal{P}$ is the set of all programs j that are "nudge-eligible" (meaning $Z_{ij} = 1$ for at least one student i).

When instantiating our method, we enforce the constraint among *High* likelihood applicants only, since estimates for the adopting subset can be noisy when the subset is small; in our simulations (Section 5), we set $L = 0.8$, and in the deployment, we set $L = 0.95$. For a nudge allocation X, let $\text{Matches}_Z(X) := \sum_{i,j} Z_{ij} \mathbf{1}\{\mu_X(i) = j\}$ denote the number of desirable matches. Then, our above discussion yields the following optimization problem:

$$X^* \in \arg\max_X \quad \text{Matches}_Z(X) \tag{1}$$
$$\text{s.t.} \quad \ell_j(\mu_X) \geq L \quad \forall j \in \mathcal{P}; \qquad X_{ij} \in \{0, 1\}, \;\; X_{ij} \leq Z_{ij} \quad \forall i, j$$

[5] A student may not functionally apply to all programs on their list if they match to a higher-ranked one. Formally, ℓ_j^{high} = (# *High*-likelihood students who functionally apply and match to j) / (# *High*-likelihood students who functionally apply to j), with $\ell_j^{\text{high}} := 1$ if there are no such applicants, and analogously for ℓ_j^{nudged}. Note that other choices of ℓ are also compatible with our approach, e.g., considering acceptance rates among all applicants.

However, we cannot directly optimize (1) in deployment: predicting $\mathbf{1}\{\mu_X(i) = j\}$ and $\ell_j(\mu_X)$ for a potential nudge allocation X requires detailed counterfactual information on each applicant's behavior and the resulting match outcome. In other words, the relationships between a nudge allocation, the objective, and the congestion constraints are intermediated by unknown applicant behavior $f(\cdot)$ and the match algorithm. Instead, we develop an approach to calculating a deployment-time allocation that approximately solves (1) in practice.

3.3 Methodological Approach and Theoretical Results

Our method, summarized in Figure 3, separates recommendation into two stages. First, using historical data, we calculate *nudge capacities* $C \in \mathbb{N}^{|\mathcal{P}|}$, the maximum number of applicants we can nudge to each program such that we expect to satisfy the congestion constraints $\ell_j \geq L$. At deployment time, we then allocate nudges given these pre-computed capacities, using an integer program whose objective function is concave increasing in the number of nudges sent to each applicant so as to broadly distribute the available nudges. This decomposition allows us to leverage historical data to ensure we satisfy the congestion constraints, and then allocate recommendations at deployment-time without advance knowledge of applicants or their preferences.

A central challenge our method solves is determining how high nudge capacities can be while satisfying the congestion constraints. We do so using the following bilevel optimization approach. Let CapacitatedNudgeAllocation, $\mathbf{CNA}(C)$, denote the nudge allocation returned by our integer program, given nudge capacities C. To evaluate candidate nudge capacities C, we make an *adoption* assumption about how applicants who *would have* submitted a ranking $\succ$ without nudges instead submit a ranking $\succ'$ with given a set of nudges. We further assume that, given any application behavior, we can evaluate whether the congestion constraints are satisfied (e.g., with a simulator). Our method for calculating nudge capacities, NudgeCapComputation, then uses **CNA** as a subroutine to evaluate candidate nudge capacities and search for the largest nudge capacity vector such that the nudge-induced applications satisfy the congestion constraints.

Theoretical result. We show that our algorithmic approach yields a set of nudge capacities and allocations that maximize a lower bound of the objective in (1) while satisfying the constraints. For theoretical simplicity, we assume the following adoption behavior.

Assumption 1 (Adoption function). An applicant receiving k nudges adopts nudges by adding one nudged program to the top of their list with probability $q(k)$, where $q(0) = 0$ and q is increasing and discretely concave. Applicants do not apply to desirable matches without adopting a nudge.

Our guarantees are stated in terms of expectations over random application and adoption behavior and the resulting match outcomes. (We also note that, in reality, applications often include desirable matches even without nudges.)

Definition 2 (Constraint Oracle and Safe recommendations). Given a nudge allocation X, let $O_j(X) \in \{0, 1\}$ be a deterministic constraint *oracle* that indicates whether the matching process induced by X satisfies the congestion constraint at program j in expectation: i.e., whether the expected number of *High* likelihood and nudged applicants who are admitted is at least a fraction L of the expected number that effectively applies.

Allocation X is safe if $O_j(X) = 1$ for every $j \in \mathcal{P}$, i.e., the constraint oracle is satisfied. Nudge capacities C are safe if the induced allocation $\mathbf{CNA}(C)$ is safe.[6]

[6]We define safety for a nudge capacity vector C with respect to a specific allocation it induces. This allows us to efficiently instantiate an oracle O, as described in Section 3.4. We note that our theoretical results can extend beyond this definition (e.g., if the oracle is able to consider arbitrary allocations compatible with a single nudge capacity vector).

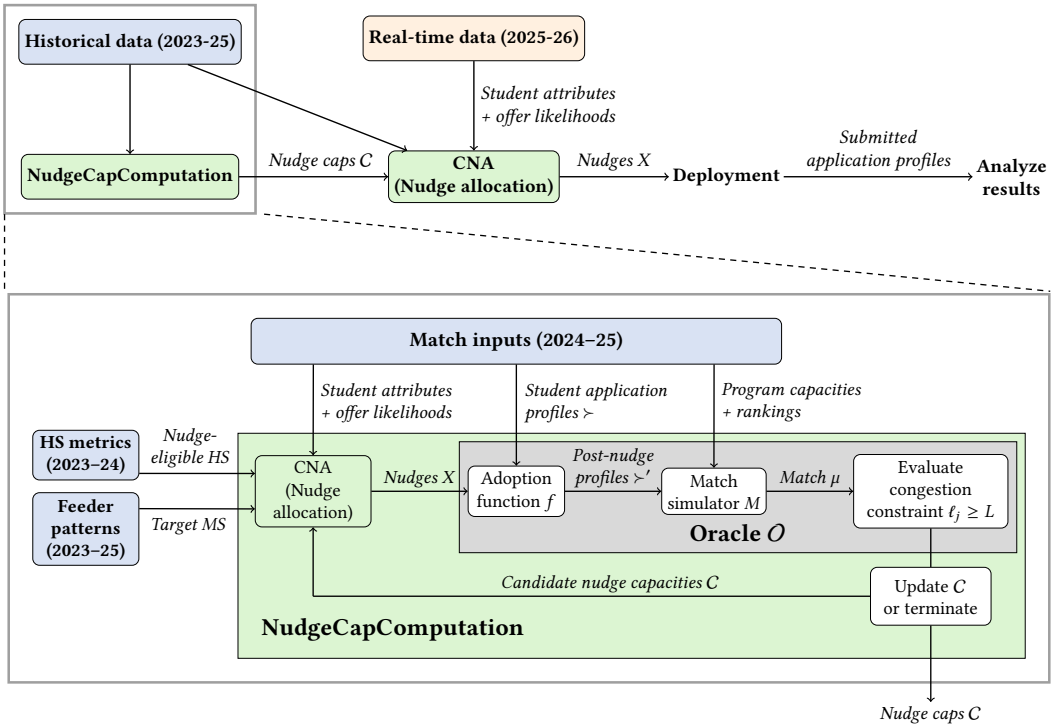


Fig. 3. **Overview of method and input data.** NudgeCapComputation computes program-level nudge capacities C using historical data (blue). Then at deployment time, we allocate nudges to new students (orange) subject to these nudge capacities. The inset shows more detail on how NudgeCapComputation searches over candidate nudge capacities C: for each candidate C, it uses CNA as a subroutine to allocate nudges, simulates applicant behavior and the resulting match outcome given the nudges, and evaluates whether the congestion constraints are satisfied.

The oracle captures how applicants adopt nudges and how their resulting applications are processed by deferred acceptance; in our setting, we instantiate O using historical data and a high-fidelity simulator of the match process.

Assumption 2 (Feasible and monotone congestion constraint). The congestion constraints are satisfied without nudges: $O_j(\mathbf{0}) = 1$ for all j. Furthermore, congestion is monotonic in nudge capacities: increasing a program's nudge capacity cannot decrease congestion at that program and cannot increase congestion at any other program. That is, holding C_{-j} fixed, $O_j\big(\mathbf{CNA}(C_{-j}, c)\big)$ is non-increasing in c, and $O_k\big(\mathbf{CNA}(C_{-j}, c)\big)$ is non-decreasing in c for all $k \neq j$.

Then our approach provides an L-approximation to the objective function in (1), among safe nudge capacities C and feasible allocations X given those capacities.[7]

[7]By considering allocations feasible under some nudge capacity vector C for which $\mathbf{CNA}(C)$ is safe, rather than over all allocations that are themselves safe, the benchmark on the right-hand side differs from our original optimization problem (1). These two sets need not coincide, depending on the oracle used to determine safety. This gap is a consequence of our approach to first calculate nudge capacities, and then optimal allocations given those capacities.

Theorem 3.1 (Approach approximately maximizes expected matches). *Suppose Assumptions 1 and 2. Let C^* be returned by NudgeCapComputation and $X^* = \textbf{CNA}(C^*)$. Then, X^* is safe and*

$$\mathbb{E}\left[\text{Matches}_Z(X^*)\right] \geq L \left[\max_{\substack{C' \text{ safe} \\ X \text{ feasible for } Z, C'}} \mathbb{E}\left[\text{Matches}_Z(X)\right] \right].$$

We now detail each component and prove this result.

3.3.1 Allocating nudges subject to nudge capacities. First, we describe how to compute an allocation X given program nudge capacities C, students $\mathcal{S}$, and desirable matches Z. We use this as a subroutine when calculating the nudge capacities themselves. Let $k_i(X) = \sum_j X_{ij}$ be the number of nudges allocated to student i under allocation X. For some utility function $u : \mathbb{N} \to \mathbb{R}_{\geq 0}$, we formulate the following integer program, denoted Capacitated Nudge Allocation (**CNA**):

$$\textbf{CNA}(C, u) : \arg\max_X \sum_{i \in \mathcal{S}} u(k_i(X))$$
$$\text{s.t.} \quad \sum_{i \in \mathcal{S}} X_{ij} \leq C_j \quad \forall j \in \mathcal{P}; \qquad X_{ij} \in \{0, 1\}, \;\; X_{ij} \leq Z_{ij} \quad \forall i, j$$

The first constraint ensures we allocate at most C_j nudges to program j. The objective maximizes total utility of the allocated nudges, for some function u that only depends on the *number* of programs student i is nudged to.

When u is linear ($u(x) = x$), **CNA** is equivalent to maximizing the total number of nudges allocated. We choose u to be strictly discretely concave and increasing, to select for a more balanced distribution of nudges. Intuitively, a balanced distribution of nudges may be both more fair and yield more desirable matches. Concentrating nudges on students who are eligible for many programs (those with greater access) may leave others with no nudges; given the geographic disparities in access throughout the city (see Figure 9a), this could perpetuate existing disparities. Further, because each student ultimately matches (and enrolls) to at most one program, allocating many nudges to the same student may have diminishing returns. We now formalize these intuitions.

Given Assumption 1 and nudge capacities C, **CNA** approximately optimizes our match objective.

Theorem 3.2 (CNA maximizes a lower bound on matches). *Suppose Assumption 1. If allocation X^* is safe and maximizes $\textbf{CNA}(C, u)$ for any increasing and strictly discretely concave u, then*

$$\mathbb{E}\left[\text{Matches}_Z(X^*)\right] \geq L \left[\max_{X \text{ feasible for } Z, C} \mathbb{E}\left[\text{Matches}_Z(X)\right] \right],$$

*i.e., **CNA** provides an L-approximation to our match objective function among feasible allocations.*

The proof in Appendix B.1 shows that under any strictly discretely concave and increasing u, **CNA** produces an allocation with *prefix-maximal coverage* (it maximizes the number of applicants who receive at least one nudge (Cov_1), and then at least two, and so on.) Then, given adoption Assumption 1 and the preservation of offer likelihoods, nudge coverage implies match guarantees: notably, preserving offer likelihoods is key to intervention effectiveness. The first part of the proof proceeds by showing that the feasible allocations induce a matroid, so the greedy algorithm (which produces prefix-maximal coverage) is optimal. For intuition, we also provide a direct construction in which any allocation that does not maximize Cov_1 can be improved by reallocating nudges from those with multiple nudges to those with no nudges, increasing the concave objective and Cov_1.[8]

[8]Note that the result is trivially true if $Z_{ij} = 1$ for all i, j, since then we can arbitrarily transfer nudges to those without nudges; but for general Z we need to show transfer feasibility, which we do so via an alternating path construction.

ALGORITHM 1: NudgeCapComputation (Diagram illustration in Figure 3 inset)

Input: Constraint oracle O, nudge capacity upper bounds $\overline{c}$
Output: Program nudge capacities C
Initialize $C^{(0)} \leftarrow \overline{c}$
for $t = 1, 2, \ldots$ **do**
  $C^{(t)} \leftarrow C^{(t-1)}$
  for *each program* $j \in \mathcal{P}$ **do**
    $C_j^{(t)} \leftarrow$ largest $c \in [0, C_j^{(t-1)}]$ s.t. $O_j\Big(\mathbf{CNA}\left((C_{-j}^{(t-1)}, c)\right)\Big) = 1$,
      which can be done via binary search between 0 and $C_j^{(t-1)}$.
  if $C^{(t)} = C^{(t-1)}$ **then return** $C^{(t)}$;

The result further implies that the exact choice of u is not important: any increasing and strictly discretely concave u yields the same optimal allocation sets. Thus, we write $\mathbf{CNA}(C)$ for the allocation returned using such a u, breaking ties according to a fixed ordering over allocations.

In summary, we have described a method (**CNA**) for computing a nudge allocation X given nudge capacities C and motivated the use of a strictly concave objective.

3.3.2 Computing nudge capacities for congestion constraints. We now describe the computation of nudge capacities $C = (C_j)_{j \in \mathcal{P}}$, the maximum number of students we can nudge towards each program while maintaining acceptance rates. Algorithm 1 (NudgeCapComputation) initializes nudge capacities at their given upper bounds and iteratively decreases them to satisfy the congestion constraints $\ell_j \geq L$ for all j. To determine this feasibility, it uses **CNA** as a subroutine to allocate nudges subject to the candidate nudge capacities and queries the oracle. For each program, we decrease its nudge capacity to the largest candidate value satisfying its congestion constraint. This process repeats until all constraints are simultaneously satisfied.

We set the effective upper bounds $\overline{c} = \left(\sum_i Z_{ij}\right)_{j \in \mathcal{P}}$, since at those nudge capacities, all applicants are nudged to all programs where they are nudge-eligible; this is the largest number of students we can simulate nudging using the observed data. Let $(C_{-j},\ c)$ denote the vector C with program j's nudge capacity replaced by c. Under Assumption 2, NudgeCapComputation returns the greatest safe nudge capacities bounded by $\overline{c}$.

Theorem 3.3 (Greatest safe nudge capacities). *Suppose Assumption 2. Then NudgeCapComputation$(O, \overline{c})$ terminates using $O\big(|\mathcal{P}|\left(1 + \sum_{i,j} Z_{ij}\right) \log\left(1 + \sum_{i,j} Z_{ij}\right)\big)$ oracle calls, and the returned nudge capacities C^* are safe. Moreover, for any safe $C \leq \overline{c}$, it holds that $C \leq C^*$ coordinatewise; thus, C^* is the unique coordinatewise greatest safe vector bounded by $\overline{c}$.*

The proof is in Appendix B.2. Putting our results together proves Theorem 3.1.

Proof of Theorem 3.1. For any safe C, the bounded capacity vector $\widetilde{C}$ defined by $\widetilde{C}_j = \min\{C_j, \overline{c}_j\}$ has an identical feasible allocation set ($\mathbf{CNA}(C) = \mathbf{CNA}(\widetilde{C})$ under fixed tie-breaking), so $\widetilde{C}$ is safe. By Theorem 3.3, C^* is safe and $\widetilde{C} \leq C^*$. Thus, every allocation feasible for Z, C is also feasible for Z, C^*. The result follows from Theorem 3.2 applied at C^*. □

3.4 Implementing our method in practice

We now discuss the instantiation of our approach in practice. We use *historical data* from the 2024-25 application cycle to compute nudge capacities: for completed application cycles, we observe a submitted *ranked application profile* $\succ^i$ for each applicant i, indicating strict preferences over some subset of programs in the absence of nudges.

This allows us to estimate counterfactual applications for any given nudge allocation X—we assume an *adoption function* $f(\succ, X)$ that simulates how applicants would have *updated* their profiles given a nudge allocation X and outputs a new collection of "post-nudge profiles." We instantiate the adoption function as follows: each applicant adds one randomly selected nudged program to the top of their no-nudge profile with fixed probability p (the *adoption rate*). In general, f may either encode domain knowledge or be learned from data on prior interventions.

We can then simulate match outcomes $\mu_X = M(f(\succ, X))$ for these post-nudge profiles, by running deferred acceptance using known program seat capacities and preferences.[9] Using this data, we instantiate the constraint oracle as

$$O_j(X) = \mathbf{1}\left\{\ell_j^{\text{high}}(M(f(\succ, X))) \geq L\right\},$$

i.e., given an allocation, we simulate adoption and the match, and then evaluate the constraint. (We just use ℓ_j^{high} in our instantiation, since ℓ_j^{nudged} becomes noisy with small nudged subsets at some programs; we verify that the constraint for the final nudge capacities generally holds among them as well.) Then, in deployment, we generate nudge allocations for observed applicants using these pre-computed capacities C (appropriately scaled and modified for robustness) and solve $\mathbf{CNA}(C, u(x) = \sqrt{x})$.[10] Recall that the exact choice of u does not matter, by Theorem B.1. This full procedure is illustrated in Figure 3.

Finally, we note that our instantiated constraint oracle is stochastic (the adoption function encodes random behavior, and multiple optimal allocations may exist)—it calculates a single draw of the congestion constraint as opposed to its expectation, meaning Theorem 3.3 does not apply as-is. (One approach is to turn Theorem 3.3 into one with probabilistic guarantees and rerun the simulation multiple times per iteration, but this prohibitively increases runtime by requiring many deferred acceptance simulations for each iteration.) To accommodate this stochasticity without substantially increasing runtime, we modify NudgeCapComputation in our implementation as follows: in each iteration, we loop through programs $\{j\}$; for each j, we simulate the match once to get an empirical estimate of ℓ_j. Then we increase j's nudge capacity if its empirical congestion metric is above $L + \epsilon$, and decrease it if it is below L. We continue until there is an iteration with no capacity changes. This modification reduces per-iteration runtime at the cost of "backtracking;" we may need to increase a program's nudge capacity, unlike in Algorithm 1. We find that this approach converges efficiently in practice; a complete run to compute nudge capacities for 56 nudge-eligible programs on the 2024-25 cycle data takes approximately 24 hours, and in practice this could be sped up further with warm-starts and further optimizations.

3.5 Model and Methods Discussion

In our setting, "desirable matches" encode our choices of "high-performing" high school programs, target middle schools, an applicant-program level offer likelihood constraint, and a transportation distance constraint. Although defining desirable matches is inherently normative, doing so is foundational to the idea of recommender systems, and our intervention goal is aligned with past interventions. Our method itself takes these as given and is agnostic to the specific student-program pairs Z deemed desirable. Furthermore, while nudges outside of the desirable pairs ($X_{ij} > Z_{ij}$) could theoretically help to accomplish the intervention goal, we decided collectively with NYCPS

[9] Since program seat capacities and preferences remain fixed in all simulations, we omit them from the notation, even though the match outcome depends on them.

[10] **CNA** can be formulated as a lexicographic max-/min-cost flow on a bipartite graph via a student-copy expansion and is therefore solvable in polynomial time; we use Gurobi for implementation convenience and directly return its output.

stakeholders that we did not want to provide any information that could be deemed lower quality. We therefore only nudge students towards desirable matches.

We also emphasize that accurately simulating counterfactual application behavior and match outcomes is crucial for our computation of nudge capacities. Following Peng et al. [2026] and the deployed offer likelihood work of Corradini and Idoux [2025a], we only nudge applicants to programs where we can reconstruct the program's admissions ranking for *counterfactual applicants.*[11] Relatedly, if our assumed adoption function f is substantively incorrect, we would misestimate behavior and congestion. For instance, we assume that students not nudged in our intervention exhibit no behavioral change; in reality, non-nudged students may still change their behavior compared to comparable applicants in previous years, due to increased reliance on the offer likelihood estimates themselves, spillover effects, or other factors. As such, our deployment this year is conservative. In the future, these other behavioral changes could be directly incorporated into the model of f.

Our theoretical results show that our method effectively optimizes for total number of desirable matches. Some natural extensions are that we could define u_i to vary by student (e.g., heterogeneous adoption rates, or importance weights). While solutions may no longer be prefix-maximal, the greedy solution still optimizes this setup, and so related results hold.

4 DESIGN DECISIONS AND EMPIRICAL CONTEXT ON THE NYC HS MATCH

In this section, we overview our data and empirical choices relevant for both our simulations in Section 5 and deployment in Section 6. We provide further detail on the NYC high school match in Appendix D.1 and on the data in Appendix D.2. We developed the system based on analysis and simulations using historical student data from the 2024-25 cycle (72,197 students), and we deploy our system for the 2025-26 cycle. We consider schools other than the 9 Specialized High Schools, which run a parallel process [NYC Public Schools, 2026b].

We define criteria for designating "high-performing high schools" in collaboration with the NYCPS Office of School Performance, based on insights from past NYCPS surveys that families care most about school performance and safety metrics [Corradini and Idoux, 2025b]. We use public metrics reported on the 2023-24 school year, the most recent results available at deployment time.

Definition 3 (High-performing high school). A high-performing high school is a high school with 4-Year Graduation Rate $\geq$ 95% (top 30% of schools), 4-Year College and Career Readiness Rating $\geq$ 60 (top 22%), and Safety and School Climate Rating $\geq$ 3.0 (top 74%).

High-performing high schools account for approximately 16% of schools and 22% of seats. Each high school has at least one *program*; different programs at the same school may provide different concentrations (such as arts), and students apply to and are admitted at the *program level.* Since not all programs have offer likelihood estimates available (e.g., auditions or essays are part of the admissions process), we further define a subset of programs at high-performing high schools that we can actually nudge select students towards, based on the availability of offer likelihoods.

Definition 4 (Potential nudge-eligible high school program). A potential nudge-eligible high school program is one that (1) is at a high-performing high school and (2) has offer likelihood estimates.

There are 56 potential nudge-eligible high school programs, across 40 schools and 17 (of 32) districts. This is the set of programs we use for simulations in Section 5. For the deployment, we requested approval from superintendents to include their schools in the intervention, and further reduced programs for a conservative deployment. Ultimately, 15 programs across 12 schools and 9 districts were included in the intervention as *deployed nudge-eligible high school programs.* Notably,

[11]The nudges do not introduce any new applicants to any *other* programs, so historical rankings over applying students are sufficient for simulating counterfactual match outcomes.

this set did not include any programs in the Bronx or Staten Island (so while we discuss effects on Bronx Latin in our simulations, it does not appear in our deployment set).

We next define criteria for target middle schools: those that do not typically send students to high-performing high schools. We target students from these middle schools in our intervention.

Definition 5 (Target middle school). A target middle school is a public NYC middle school (excluding charter schools) that sent less than 15% of its students to high-performing high schools in the 2024-25 cycle, and averaged across 2023-24 and 2024-25 cycles. Note that we consider all high-performing high schools here, not just nudge-eligible ones.

This resulted in 240 target middle schools. This is the set of target middle schools used in all simulations. Demographically (all numbers based on the 2024-25 cycle), 83% of students at these schools are eligible for Free/Reduced Lunch (FRL) (versus 73% overall). 11.5% of students at high-performing high schools came from target middle schools, even though target middle schools account for 33.6% of incoming high school students and over 50% of students at target middle schools have *High* offer likelihood to at least one nudge-eligible program within 45 minutes of their home. For the deployment, due to changes in schools still open and other related rationales, we had a total of 218 target middle schools. Further, nudges were sent in two batches, and we dropped a set of middle schools from Batch 2 in response to stakeholder feedback after Batch 1. This criterion was applied to both treatment (25 dropped) and control (22 dropped).

Finally, a student is eligible to receive emails encouraging them to apply to a nudge-eligible high school program if they meet the following criteria, determined in collaboration with many NYCPS offices and informed by parent focus groups held in Summer 2025.

Definition 6 (Nudge-eligible student for a program). Any student i is "nudge-eligible for a program j" if they attend a target middle school, their middle school sent less than 2% of its students to program j in the 2024-25 cycle, they have a *High* chance of receiving an offer to program j, and their home census tract's centroid is within 45 minutes of program j's high school.

We use OpenTripPlanner to estimate travel time using any combination of subway, bus, and walking. In data from the 2024-25 cycle, using zip code as an approximation for census tract, this led to about 11,500 total students who are nudge-eligible for at least one program. 88% of these students are FRL-eligible, compared to 73% in the general student population and 83% at target middle schools.[12] Geographic distributions of these schools and students are shown in Figure 8.

5 RESULTS: CONGESTION IN SIMULATION

Here we demonstrate the implications of congestion and evaluate our approach using simulations on 2024-25 cycle data. This analysis is "in distribution": we compute nudge capacities (using our approach in Section 3) and then evaluate match outcomes *as if we nudged applicants in that same cycle*, under various nudge allocation methods and adoption assumptions.[13] To account for stochasticity in nudge allocations and adoption, we report mean values over repeated simulations.

5.1 Empirical setup

We use the definitions of nudge-eligible high schools and students specified in Section 4. Notably, for all simulations in this section we use the "full" set of 56 potential nudge-eligible high school programs, including programs that we ultimately cut in the deployment (e.g., Bronx Latin). We allocate nudges among nudge-eligible students according to each of the methods defined below. At

[12] If we filter for programs within 30 minutes of a student's home instead of 45, the number of students nudge-eligible for at least one program becomes about 6,500. Among these students, 87% are FRL-eligible.

[13] In contrast, the deployment is "out of distribution," run on 2025-26 application data using capacities computed for 2024-25.

various adoption rates, we simulate modifications to their true submitted application profiles (f in our above notation) and the resulting match outcomes. All non-nudge-eligible students retain their original application profiles. For each adoption rate p, we compute and enforce nudge capacities to target acceptance rates among *High* offer likelihood students (ℓ_j) above 80% at p% adoption.[14] We compare three methods, calculating nudge capacities as defined in Section 3:

- **Naive**: Nudge *all* nudge-eligible student-program pairs, without using nudge capacities.
- **Capacitated**: Allocate nudges subject to nudge capacities, maximizing total number of nudges (linear objective), i.e., $\mathbf{CNA}(C, u(x) = x)$.
- **Balanced Capacitated**: Allocate nudges subject to nudge capacities, maximizing a concave utility of nudges, i.e., $\mathbf{CNA}(C, u(x) = \sqrt{x})$.[15]

Balanced Capacitated allocation is our proposed method, while the others disambiguate the effects of nudge capacities and of the concave objective function. In Appendix C.1, we further analyze *non-personalized* recommendations in simulation, in which all students from target middle schools are nudged to all nearby nudge-eligible programs (reflecting the design of prior interventions).

5.2 Results

Nudge capacities preserve program-level acceptance rates. Figure 1b shows that the nudge capacities indeed decrease program-level congestion, retaining acceptance rates for *High* offer likelihood students above the targeted 80% threshold. This improves upon results from **Naive** allocation, where acceptance rates among *High* offer likelihood students fall below 80% for 7 of 56 programs at 20% nudge adoption, and for 15 of 56 programs at 50% nudge adoption. In Appendix C.1, we show that even **Naive** personalized recommendation substantially outperforms non-personalized recommendation, in which most programs reject more than half of applicants even at 20% adoption.

Nudge capacities address disparities in match rates to nudged programs. Figure 2a depicts match rates to nudged programs by zip code. Nudge capacities lead to increased match rates among nudged students across the city, addressing geographic disparities that arose with the **Naive** approach.

Figure 2b shows that nudge capacities also address disparities in match rates by level of student access. With **Naive** nudge allocation, match rates among students with access to only one nudge-eligible program are much lower (despite all students adopting only *one* nudged program). Enforcing nudge capacities with our **Balanced Capacitated** method leads to increased match rates across students of all access levels, and particularly large gains among students with limited access to nudge-eligible programs (increasing match rates for students with 1 inventory program from 51.6% to 76.2%, and those for students with 5+ inventory programs from 91.3% to 95.8%).

Balanced Capacitated *allocation increases match rates while approximately preserving the total number of students who match to nudged programs.* Notably, while the number of nudges and nudged students decreases when imposing capacities (leading to higher match rates), using our method retains the number of *matches to nudged programs*—"desirable matches"—which our intervention goal is to maximize. In our simulations at 20% (Table 1), our method actually achieves *more* matches to nudge-eligible programs than the **Naive** method of allocating nudges to all eligible students (which has substantially lower match rates). Thus, nudge capacities can help address congestion without sacrificing overall effectiveness of the intervention. Intuitively, the constrained nudges steer applicants toward the programs they are more likely to be admitted to in equilibrium.

[14] Note that in the deployment, we ultimately targeted 95% acceptance rates.

[15] As shown in Theorem B.1, any increasing, strictly concave function has the same set of optimal allocations. We rely on Gurobi's returned solution to tiebreak among the optimal set. Note also that solutions maximizing a concave objective are also optimal for the linear objective.

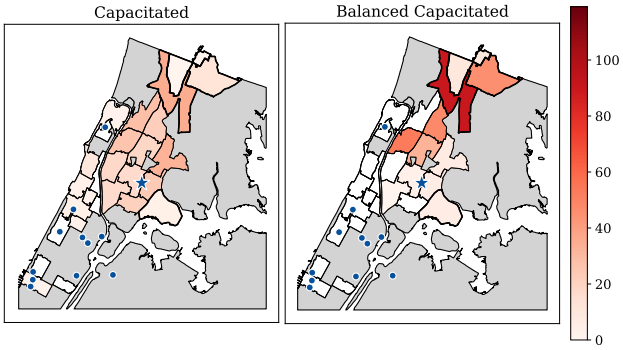


Fig. 4. **Simulated geographic shift in students nudged to Bronx Latin at 20% adoption.** Geographic differences between recommendation methods in simulation. The shade of red reflects the number of students from each zip code nudged to Bronx Latin under each allocation method. Nudge-eligible programs are indicated by the blue star (Bronx Latin) and dots (other). Balanced allocation gives nudges to students in zip codes *without* access to any other programs (indicated with thicker outline).

Allocation with concave objective leads to a geographic shift in prioritized students. Tables 1 and 2 also reveal that **Balanced Capacitated** allocation improves upon **Capacitated** in number of unique students nudged (Cov_1) and matched, empirically supporting the theoretical claims in Section 3.3.1. To illustrate the key difference between the allocation methods, Figure 4 depicts the distribution of students nudged to Bronx Latin in the simulation, the only nudge-eligible school in the Bronx. Due to geographic constraints, many students (1304) have Bronx Latin as their *only* nudge-eligible program. Of these students, 630 reside in zip codes where *no* student has access to *any other* nudge-eligible programs. Under **Capacitated** allocation, nudges to Bronx Latin are distributed arbitrarily among eligible students. In contrast, the **Balanced Capacitated** allocation prioritizes students without access to other programs, shifting more nudges towards students in the Upper Bronx (especially those in zip codes with no access to other nudge-eligible programs).

This shift leads to meaningful changes in match outcomes. As depicted in Figure 10, **Capacitated** allocation yields 9 matches from Bronx Latin-only eligible zip codes, whereas **Balanced Capacitated** yields 25. Importantly, this does *not* come at the expense of target students elsewhere, as they still receive nudges to other programs: among students from other zip codes with access to Bronx Latin, **Capacitated** allocation yields 517 matches to other nudged programs, whereas **Balanced Capacitated** yields 611. Thus, balanced allocation increases desirable match outcomes for students with limited choice sets, *without* reducing aggregate matches among other students.

These simulations show that our method effectively achieves the intervention goal of increasing matches to high-performing programs while decreasing disparities in acceptance rates.

6 DEPLOYMENT AND RANDOMIZED CONTROLLED TRIAL

In Fall 2025 (2025-26 admissions cycle), we deployed our method in a personalized intervention in collaboration with NYCPS. We sent emails to a subset of nudge-eligible students (our treatment group), highlighting high-performing programs that might be a good fit for them. We ran a pre-registered randomized controlled trial to evaluate the impact of our intervention.[16] Discussions

[16]Pre-registered at the AEA RCT Registry: https://www.socialscienceregistry.org/trials/17146

Allocation Method	Total # Nudges	# Students Nudged	Match Rate	# Matched to Nudged	# Matched to any Nudge-eligible
Naive	26,325	11,417	0.70	1,597	1,989
Capacitated	16,175	7,838	0.98	1,520	1,947
Balanced Capacitated	16,175	9,804	0.89	1,744	2,153

Table 1. **Simulation results at 20% student adoption.** While the **Naive** method results in more total nudges and total students nudged, recommendation-induced congestion leads to low *match rates* to nudged programs (among students who effectively apply to a nudged program, the fraction that is accepted). Enforcing nudge capacities results in higher match rates to nudged programs; on top of this, our **Balanced Capacitated** method still results in *more* students matched to nudged programs and nudge-eligible programs overall (*desirable matches*).

with NYCPS stakeholders guided our congestion constraint policy decisions: in particular, we aimed for at least 95% acceptance among *High* likelihood applicants, at adoption rates of up to 40%. These decisions were conservative, to account for uncertainty over student adoption (Figure 11 illustrates the tradeoff between intervention effectiveness and potential congestion effects from computing nudge capacities under incorrect assumptions about adoption; in this first deployment we conservatively err on the side of preserving *High* offer likelihood acceptance rates).

All students—including control students and those not eligible for the intervention—can access personalized offer likelihood estimates and program metrics through other sources (e.g., MySchools, commute time calculators). Thus, our intervention does not make available any *new* information to treatment applicants, compared to others; rather, it highlights existing information in a way that encourages applicants to consider applying to high-performing programs that may be a good fit. Furthermore, applicants have ultimate agency in their application decisions and can ignore nudges if the programs do not align with their true preferences.

6.1 Experiment design

Nudge-eligible High School Programs. As mentioned in Section 4, our deployment set of nudge-eligible programs was a subset of the potential nudge-eligible programs, resulting in 15 total programs. It did not include any programs in the Bronx or Staten Island.

Experimental Population. Nudge emails were sent in two batches during the application cycle. For each batch, we pulled data 1–2 days prior to email delivery and determined the experimental population for that batch as being nudge-eligible students (as defined in Section 4) whose applications were *in-progress* (they had opened their application, meaning we had offer likelihood estimates for them, but they had not yet submitted). Offer likelihood predictions are dependent on a student's lottery number, which is only generated when they begin their application in the online portal. Thus, this design allows us to balance reaching more students and sending them communications earlier in their application process (hopefully before they have finalized their decisions).

Randomization and Treatment Assignment. We divided our experimental population into two groups: control students received all standard NYCPS communications, and treatment students received the standard communications plus *nudge emails* with program recommendations. We randomized at the middle school level, assigning each target middle school to 0, 50, or 100% treatment intensity prior to the first email batch. This design allows for the possibility of measuring spillovers across students who attend the same middle school. The number of schools assigned to each treatment intensity (keeping 50% and 100% treatment intensity groups at equal size) was

chosen such that—based on last year's number of nudge-eligible students per middle school—the expected total number of treatment applicants would be approximately 500. We assigned middle schools to treatment intensities via stratified sampling, with strata defined by borough and expected number of nudge-eligible applicants.

Individual student treatment assignment was implemented separately for each batch. For each batch, we assigned students to treatment or control within each school according to the school's pre-assigned treatment intensity (0%, 50%, or 100%), by randomly sampling the appropriate fraction of students from the experimental population at that middle school. Balance tests in Appendix D.3 support that the randomization was effective; we further prespecified and include stratum fixed effects in our primary analyses, to account for realized differences in the group samples.

Nudge Allocation for Treated Applicants. Among treatment students in each batch, we determined which nudge-eligible high school programs to recommend to which students using the integer program described in Section 3.3.1. Each program's nudge capacity was pre-computed as described in Section 3.3.2 using data from the 2024-25 cycle, targeting an acceptance rate of at least 95% among high likelihood applicants at up to 40% student adoption. Because nudges were sent in two batches, we solved the optimization problem separately for each batch. Nudge capacities were scaled according to the batch size: if Batch 1 contained x% of the expected total of 500 treatment students, we would scale the Batch 1 nudge capacity for each program by x% and Batch 2 nudge capacities by (100-x)%. Ultimately, with 353 treatment students in Batch 1 and an overall target of 500, each program's nudge capacity in Batch 1 was approximately 70.6% of its overall nudge capacity.

Defining Counterfactual Nudges for Control Applicants. For the control group, the set of programs they *would have been nudged to* is not clearly defined. There is no clear counterfactual for what would have happened if each student had instead been in the treatment group, since nudge allocations depend on the composition of the treatment group itself. For example, we cannot simply add each control group student (individually) to the treatment group and re-solve the optimization: due to the clustered sampling at the middle school level, adding each control student individually would not reflect the true counterfactual scenario where their middle school had been assigned to treatment; since congestion patterns are geographically correlated, this would under-estimate their expected congestion, and control group students would counterfactually receive larger nudge lists.

To address this challenge, we take advantage of the fact that, by design, the control and treatment groups are distributionally identical in expectation. For each batch, we therefore solve a separate, parallel optimization problem for the control group, as if the control and treatment groups had been swapped entirely. This produces, for each control applicant, a counterfactual set of programs to which they would have been nudged had they instead been assigned to treatment.

Because the control group was larger than the treatment group, we proportionally scaled program capacities in the control optimization. For each batch, if the control group had x% as many students as the treatment group, we scaled each program's nudge capacity by x%. Using the resulting counterfactual nudge sets, we can directly compare behavior and match outcomes between groups.

Nudge Emails. The 2025-26 high school application opened on October 7 and closed on December 4, 2025. We sent nudge emails from the NYCPS High School Enrollment email account[17] on October 30 (Batch 1) and November 19 (Batch 2). The email template is in Figure 15. Our emails explained to applicants that they had a *High* chance of offer at high-performing high school programs within 45 minutes of their home and listed these programs with tracked links to the school webpages

[17]This NYCPS email account also sent other application-related emails to all students during the application period, such as reminders to submit applications before the deadline.

(allowing us to observe clicks for each school, but not whether an applicant opened the email). On November 20, we sent targeted follow-up emails to students from Batch 1, personalized based on whether their saved application profile (if any) included no high-performing programs, high-performing programs but ranked below a program *not* on the high-performing programs list where they had *High* offer likelihood, or at least one high-performing program at the top of their list.

Operational details. Deploying the intervention in collaboration with NYCPS involved many stakeholders and coordinated entities. The intervention and all family-facing text were approved by both University and NYCPS IRBs. Consistent with our data use and non-disclosure agreements, no individual-level identifiers were shared with University partners outside of secure NYCPS systems.

Implementation choices were shaped by both technical and operational constraints, and key design decisions were made in collaboration with multiple NYCPS offices. Additionally, we conducted six virtual focus groups with 8th- and 9th-grade families in English and Spanish, helping to inform decisions such as what constituted a reasonable distance from a student's home. The email text was drafted and edited in coordination with multiple NYCPS offices and aligned with the High School Admissions Team's existing communications language and schedule. Consistent with NYCPS norms, emails were sent to the *parents or guardians* of applicants. Personalized emails were translated into 12 languages by the NYCPS translations team and then sent from the High School Enrollment email address, potentially in multiple languages as appropriate for the recipient. While we considered integrating nudges into the application portal, sending text messages, and mailing physical letters, these options were not feasible this year and we delivered nudges solely via email.

All major project decisions were discussed with and approved by NYCPS leadership, including the Chief Enrollment Officer, Deputy Chancellor of School Leadership, Acting First Deputy Chancellor, and Chancellor. We obtained approval from superintendents of all deployment high schools and informed staff at treated middle schools about the intervention and which students received nudges to which programs. Based on stakeholder feedback after the first batch, we removed students from several middle schools for the first batch's reminder emails and the second batch.

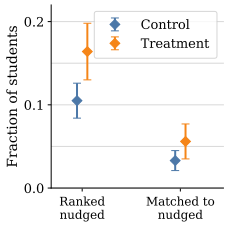
After the application deadline but before match outcomes were released to applicants, we also administered an optional post-application survey to gather descriptive evidence on families' perceptions of the emails and demand for recommendations. We sent a post-application survey to approximately 2,200 applicants with a $20 incentive. Survey responses were not used as primary experimental outcomes.

6.2 Experiment results

We ended up with 822 control and 444 treatment students. Here, we report results on treatment effects (application behavior and match outcomes), congestion, and general family engagement and perceptions, with supporting results in the appendix.

Treatment effects on application rates and match outcomes. Figure 5 contains the primary experimental results. Among treatment applicants, 16.4% ranked at least one program they were nudged to at least somewhere on their list; among control applicants, 10.5% ranked a program they would have been nudged to. In terms of match outcomes, 5.6% of treatment applicants matched to a nudged program, and 3.3% of control applicants matched to a program they would have been nudged to. *No* treatment students were rejected from a program they were nudged to (out of 25 effective applicants, i.e., not counting applicants who listed a nudged program but were accepted to another program they ranked higher), and 1 control student was rejected from a program they would have been nudged to (out of 28 effective applicants), for an overall 98% acceptance rate.

We fit pre-registered logistic regression models with standard errors clustered at the middle school level. Table 5b reports estimates of the treatment effect on application behavior, both with



(a) **Outcome rates.**

Outcome	Control rate	Treatment rate	Stratum FE	Treatment coef.	p-value
Ranked nudged	0.105	0.164	Yes	0.561 (0.219)	0.011**
			No	0.521 (0.232)	0.024**
Matched to nudged	0.033	0.056	Yes	0.643 (0.356)	0.071*
			No	0.564 (0.346)	0.104

(b) **Regression-estimated treatment effects (Logit): primary outcomes.** 1,266 total observations. SEs in parentheses. * $p < 0.10$, ** $p < 0.05$.

Fig. 5. **Application behavior by treatment condition.** Regressions are fit with robust standard errors clustered at the school level, estimated with and without stratum fixed effects (borough and school size). CIs in the figure are 95% CIs.

and without stratum fixed effects. In the pre-registered model with stratum fixed effects, treatment significantly increased the likelihood of ranking a recommended program ($\hat{\beta}$=0.561, SE=0.219, p=0.011). For interpretability, we compute average marginal effects to translate this estimate to the probability scale, finding that treatment increased the probability of ranking a recommended program by 6.5 percentage points (SE=2.6pp, p=0.013), relative to a control mean of 10.5%. There is also suggestive evidence that the treatment increased matches to recommended programs ($\hat{\beta}$=0.643, SE=0.356, p=0.071). On the probability scale, this corresponds to a 2.7 percentage-point increase (SE=1.6pp, p=0.103), relative to a control mean of 3.3%. In terms of *where* in a student's application profile these nudged programs were listed, control and treatment are distributionally similar: of students who ranked a nudged program (in control and treatment groups, respectively), 22.1% and 23.3% listed it first, and 47.7% and 47.9% in the top 3.

Next, we consider applications to all nudge-eligible programs, not just those the applicant was actually nudge-eligible for or nudged to. These outcomes are less dependent on our empirical choices, such as the calculation of hypothetical nudges for the control group. Generally, the additive effect sizes remain unchanged: 36.5% of treatment applicants ranked at least one nudge-eligible program somewhere on their list, versus 29.6% of control applicants. 8.3% of treatment applicants matched to a nudge-eligible program, versus 6.3% of control applicants.

We run robustness checks with linear probability models, alternative fixed effects specifications, and additional outcome specifications; these do not meaningfully change the results (Appendix D.4, D.6). We also report results from additional analyses, including for spillover effects, local average treatment effects, and heterogeneous treatment effects (Appendix D.6).

Link clicks. The above analyses follow an intent-to-treat design: we include *all* applicants assigned to treatment or control. For insight on engagement, 88 treatment students (19.8%) clicked on the trackable links, and among those, 25 (28.4%) ultimately ranked a nudged program (compared to 16.4% of all treatment applicants). We do not observe email open events.

Post-application survey responses. We received 239 responses (42 treatment, 61 control, 136 from non-target middle schools). 25/42 treatment families reported *reading* a nudge email; among those, 22/25 (88%) reported that the email was helpful and 11/25 (44%) reported that it affected their decisions. 173/195 (89%) of non-treatment families reported that they would have liked to receive program recommendations (2 families did not answer the question and are omitted). Although these responses should be interpreted descriptively due to potential selection bias, they suggest substantial perceived value of personalized recommendations.

The survey also shed light on why some families did *not* add nudged programs to their application; the most common reported reason was that the recommended programs' academic interest area was not aligned with the student's interests, and many families stated that extracurricular offerings would be an important factor for them in recommended programs. This suggests that while offer likelihood and commute distance are important dimensions for identifying potentially relevant programs, future recommendation systems may benefit from incorporating further personalization regarding students' academic and extracurricular preferences.

6.3 Scale-up and robustness: What are the limits of student-side interventions?

Our experiment design was intentionally conservative to account for uncertainty in the first deployment: we reduced the number of nudge-eligible programs, determined nudge capacities assuming an adoption rate of 40%, and limited the expected treatment group size to 500. These precautions reflect NYCPS' real concerns about recommendation-induced congestion and its effect on offer likelihoods; they provided safeguards against unexpectedly high adoption, year-to-year shifts in application behavior, or other deviations from our modeling assumptions.

Having deployed the intervention and observed its effects in practice, we may wish to relax these constraints and "scale up" such an intervention to increase its reach in the future. Along these lines, we explore two questions: (1) What is the maximum impact the modeled intervention could have, without increasing the supply of high-performing programs? (2) How robust are our congestion calculations to year-to-year distribution shifts in the market environment and applicant behavior?

Scale-up. Here, we ask: how impactful can demand-side interventions be? This question informs the necessity of supply-side interventions or other structural changes. We answer this question by calculating nudge capacities using our framework assuming 100% adoption—since the capacities preserve acceptance rates, this calculation approximates how many applicants can simultaneously, safely, apply and match to high-performing options. Under our model, any larger effect, from demand-side behavior changes alone, must displace existing students with *High* offer likelihoods.

Using the 2024-25 applicant population, we compute nudge capacities to maintain the targeted 95% acceptance rate among *High* offer likelihood students at 100% adoption. We use the set of 15 *deployed* high school programs. We run simulations akin to those in Section 5 to assess the maximum number of students we could have nudged and that would have matched to nudge-eligible programs. With the set of 15 deployed programs, this procedure yields 873 desirable matches for the 4,982 nudge-eligible students (in reality, 42 such matches occurred that year).

There are three further paths to scaling up the size of future interventions. First, our definition of high-performing schools was stricter than in previous interventions; relaxing this, potentially with multiple categories of schools, could allow including more schools. Relatedly, as the survey responses suggest, applicants also have (likely heterogeneous) preferences for other school characteristics, such as extracurriculars offered; incorporating such personal preferences in recommendations could further reduce congestion. Finally, we have so far considered the offer likelihood design as fixed; it may be possible to jointly design the intervention and likelihood calculations, such as by incorporating intervention effects into the estimators. These changes would expand the set of feasible intervention designs.

Distribution shift. Each year, the distribution of applications across programs may change. For instance, offer likelihood estimates themselves (especially as they are more widely advertised and trusted), other policy decisions, demographic shifts, and program changes could change applicant behavior over time. Given our use of historical data in computing nudge capacities, it is important to understand robustness to these year-to-year distribution shifts.

We re-compute nudge capacities using the (deployment year) 2025-26 applicant population, targeting the same 95% acceptance rate among *High* offer likelihood students at 40% adoption,[18] and compare these values with nudge capacities computed using the same parameters on the 2024-25 applicant population.[19] We use our 2024-25 match simulator (using that cycle's seat group targets, rule sets, etc.[20]) for both calculations, to focus on changes from applicant behavior, as well as to reflect the reality that an accurate match simulator for an admission year may not be available until after nudges must be computed.

This comparison reveals two forms of year-to-year shift. First, the market environment itself changed: while the 2024-25 simulator reproduces the 2024-25 match with 98.2% accuracy, it achieves only 88.6% accuracy for the 2025-26 match. This suggests that changes beyond student applications (e.g., program characteristics, policies, or data availability) can reduce the accuracy of offer likelihood and nudge capacity calculations across years.[21] Thus, while using a previous year's simulator may be necessary given the data available at deployment time, these estimates necessarily inherit errors from shifts in the market environment. Second, we compare computed nudge capacities between years (Figure 14a)—since we hold the match simulator fixed, these differences in nudge capacities are only due to shifts in the observed application data. Overall, nudge capacities are strongly correlated across years (R=0.945), although we do observe notable variation at individual programs. The mean absolute difference in nudge capacities between years was 48 (the average program nudge capacity in 2025 was 144), and the mean difference from 2024-25 to 2025-26 was -16.

This aggregate decline is largely driven by 4 of the 15 programs, whose computed nudge capacities dropped to 0 when computed on 2025-26 data; collectively, the drop in nudge capacities at these programs was 246 (compared to a total net decline of 248). At these programs, there seems to have been a notable increase in applications and resulting congestion; our target acceptance rates could no longer be maintained in simulations, even without additional simulated nudges. Among the other programs, nudge capacities fluctuate in both directions but exhibit little net change.[22]

Together, these results imply that computing nudge capacities on historical data is informative even under the observed distribution shift, but that they do not induce exact guarantees. It is thus important to deploy conservative interventions, as we did—an open methodological question remains on how to best be robust to such shifts at scale. Overall, our results suggest promise in being able to effectively scale up the intervention while preserving offer likelihood constraints.

7 CONCLUSION

In this paper, we study how to design personalized recommendations in capacity-constrained matching markets. We characterize recommendation-induced congestion by showing that simultaneous adoption can sharply and heterogeneously reduce acceptance rates, failing to preserve a key

[18]While some 2025-26 applicants adopted nudges in response to our intervention, our treatment group was intentionally small. We simulate 40% adoption on *top* of the submitted profiles, making our computed capacities more conservative.

[19]At some programs it was impossible to preserve *High* offer likelihood student acceptance rates above 95% because the rate dropped below 95% in the true outcomes this year; we set these nudge capacities to 0.

[20]For programs with manual rankings, such as after auditions, we use their true rankings over the 2025-26 applicants.

[21]While it is unclear, even in hindsight, what exact changes lead to this drop, two plausible factors are: (1) year-to-year changes in program capacities, and (2) our 2025-26 data does not perfectly identify FRL status applicants, instead combining that data with other status indicators, such as whether a student is an English Language Learner.

[22]Some of this program-level variation reflects limitations of the data available for our simulations, as opposed to purely changes in underlying congestion. In particular, we only received offer likelihood data for students captured in our two intervention batches, and therefore do not observe the full set of nudge-eligible pairs for the 2025-26 application cycle. Based on past years' data, we expect that the number of nudge-eligible students at the time of the Batch 2 data pull is about *half* the total number. For 5 programs, this appears to have been the limiting factor on the maximum nudge capacity compared to the previous year.

signal for the recommendations themselves. Our results suggest that recommendation systems in capacity-constrained environments (school matching, labor markets, or dating platforms) should be designed with congestion in mind—individual-level interventions can have substantial system-wide externalities, inducing both inefficiencies and disparities. Then, we propose a generalizable method for recommendation in equilibrium, preserving offer likelihoods while effectively achieving the intervention goal of increasing the number of matches to nudged programs. Finally, we implement our approach in a large, real-world market—the NYC High School Match—and evaluate it through a randomized controlled trial. Treated students were 57% more likely to apply to a recommended program (p=0.011) and 71% more likely to match to a recommended program (p=0.071), demonstrating that recommendations shift application behavior and providing suggestive evidence of changes in matches. To our knowledge, this is the first field deployment of a recommendation system to control congestion in a centralized school matching platform.

We foresee that similar analyses and interventions would be informative in other capacity-constrained matching contexts—including labor, dating, and healthcare markets. Our view casts recommender systems as interventions that affect both individual recipients and the market as a whole. Under this view, safe, effective deployment of recommenders requires both measuring systemic effects and proactively designing for them. Our work highlights several potential design principles. While recommendations should incorporate match probabilities, they should do so in equilibrium, incorporating recommender effects themselves. This is especially important because, as we highlight, increased congestion can exacerbate existing disparities, disproportionately affecting applicants with the fewest existing opportunities. More broadly, recommendation design is policy design: the choices of "desirable matches" and the "congestion constraint" reflect decisions about how policymakers may want to change the status quo and preserve it, respectively.

Finally, our intervention is ultimately constrained by the limits of demand-side interventions in capacity-constrained settings. It cannot be scaled up indefinitely while preserving offer likelihoods, and the supply of seats at high-performing schools limits the potential increases in matches. We quantify such limits—larger changes may require supply-side interventions, such as increasing the supply of seats in high-performing schools. These considerations have analogues in other settings, in which improving recommendations complements, rather than substitutes, structural changes.

Discussion and Limitations

Our experimental design and deployment make several policy choices, in consultation with our deployment partners. First, we prioritize preserving offer likelihoods for *High* likelihood students, allowing for displacement of lower-likelihood applicants (quantified in Appendix C.2). Intuitively, this increases assortativity with respect to each program's admissions priorities, while retaining reliability of *High* offer likelihood signals. If desired, the congestion constraint can be adapted to alternative policy goals.

Additionally, showing offer likelihood predictions to students may change student behavior even without further nudge interventions, and this can shift the congestion we are trying to anticipate when designing our intervention (recall that all students see offer likelihood predictions [Corradini and Idoux, 2025a]). This is one of the reasons we were conservative in our intervention. In future deployments, explicitly modeling behavior changes in response to both offer likelihoods and recommendations could support more robust design. More broadly, *jointly* designing offer likelihood estimates alongside recommendations may further improve outcomes in equilibrium.

Our optimization framework also makes simplifying assumptions. These assumptions were made to retain model transparency and simplicity, which were especially important for our first deployment. First, we assume homogeneous nudge adoption; future versions of the system could capture heterogeneity along these lines and weight recommendations accordingly. Successfully

implementing richer personalization has the potential to further improve match outcomes. Furthermore, our method currently produces batch recommendations and relies on accurately simulating match outcomes; extending it to settings with real-time recommendations and stochastic matches is an important direction for future work. Integrating recommendations directly into the application portal may enable dynamic recommendations and improved personalization based on student preferences. However, these extensions would make implementation and analysis more complex, and would require careful validation to ensure the model does not end up systematically de-prioritizing more disadvantaged students or amplifying prediction error.

Researchers and NYCPS do not observe true applicant preferences. While our survey responses and past surveys [Corradini and Idoux, 2025b] indicate that safety, academic performance, and distance are most important to families, other factors not included in our study are also important. For example, our survey results also highlight the importance of program interest area (Appendix D.7). Prior work has shown substantial preference heterogeneity, including by demographic group [Abdulkadiroğlu et al., 2020, Burdick-Will et al., 2024, Corradini and Idoux, 2023, 2025b, Hailey, 2022, Harris and Larsen, 2015]. Relatedly, our school performance measures—though decided in conversation with NYCPS stakeholders—do not capture other measures of school quality, such as causal value-add [Angrist et al., 2024b]. However, applicants have ultimate decision-making authority; when our nudges do not match applicants' personal preferences, they may not be adopted, potentially mitigating negative effects. On the other hand, it is possible that some nudges were adopted without a full understanding of the program's offerings compared to other options. Thus, matching to a nudged program does not necessarily imply an individual welfare gain.

ACKNOWLEDGMENTS

This research is supported by the Alfred P. Sloan Foundation, a Cornell Center for Social Sciences seed grant and the center's Regulated Research Environment, and a William T. Grant Foundation Scholars Award to NG. EC is supported by NSF DGE #2139899, a Cornell Tech Digital Life Initiative doctoral fellowship, and a Siegel PiTech PhD impact fellowship. NG is supported by NSF CAREER #2339427, and Cornell Tech Urban Tech Hub, NASA, Google, Mastercard, and Amazon research awards. We are grateful to Samuel Daunt, Katrina Fortune, Allie Jackson Hing, Amy Lee, Brett Saffer, Perry Stein, Christian Barnes, Matthew Broggini, Nadiya Chadha, Yolanda Fana, Tamara Lindor, Ahmad Shilleh, Diying Wu, Trevonda Kelly, Eric Ashton, Michael Cho, Aurora Moraes, Benjamin Cosman, and Justin Matthews at NYC Public Schools for their guidance, expertise, and support throughout this project. We also thank Gabriel Agostini, Peter Bergman, Samuel Breckenridge, Christopher Campos, Sarah Cohodes, Viola Corradini, Sophie Greenwood, Clemence Idoux, Russell Legate-Yang, Irene Lo, Anushka Murthy, Parag Pathak, Kevin Ren, Ignacio Rios, Emily Ryu, and other members of the Garg Lab and Cornell AIPP for valuable conversations and feedback.

REFERENCES

Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. 2005. The New York City High School Match. *American Economic Review* 95, 2 (April 2005), 364–367. https://doi.org/10.1257/000282805774670167

Atila Abdulkadiroğlu, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. Do Parents Value School Effectiveness? *American Economic Review* 110, 5 (May 2020), 1502–1539. https://doi.org/10.1257/aer.20172040

Roy Adams, Katharine E Henry, Anirudh Sridharan, Hossein Soleimani, Andong Zhan, Nishi Rawat, Lauren Johnson, David N Hager, Sara E Cosgrove, Andrew Markowski, et al. 2022. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nature medicine* 28, 7 (2022), 1455–1460.

Gabriel Agostini, Emma Pierson, and Nikhil Garg. 2024. A Bayesian spatial model to correct under-reporting in urban crowdsourcing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 21888–21896.

Patrick Agte, Claudia Allende, Adam Kapor, Christopher Neilson, and Fernando Ochoa. 2024. *Search and Biased Beliefs in Education Markets.* Technical Report. National Bureau of Economic Research.

Narges Ahani, Paul Gölz, Ariel D Procaccia, Alexander Teytelboym, and Andrew C Trapp. 2024. Dynamic placement in refugee resettlement. *Commun. ACM* 67, 5 (2024), 99–106.

Robert Ainsworth, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola. 2020. *Information, Preferences, and Household Demand for School Value Added.* Working Paper 28267. National Bureau of Economic Research. https://doi.org/10.3386/w28267

Robert Ainsworth, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola. 2023. Why do households leave school value added on the table? The roles of information and preferences. *American Economic Review* 113, 4 (2023), 1049–1082.

Omar Al-Ubaydli, John A List, and Dana L Suskind. 2017. What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review* 107, 5 (2017), 282–286.

S Nageeb Ali and Ran I Shorrer. 2025. Hedging when applying: Simultaneous search with correlation. *American Economic Review* 115, 2 (2025), 571–598.

Claudia Allende, Francisco Gallego, and Christopher Neilson. 2019. *Approximating the equilibrium effects of informed school choice.* Technical Report.

Maxwell Allman, Itai Ashlagi, Irene Lo, Juliette Love, Katherine Mentzer, Lulabel Ruiz-Setz, and Henry O'Connell. 2022. Designing school choice for diversity in the San Francisco Unified School District. In *Proceedings of the 23rd ACM Conference on Economics and Computation*. 290–291.

Maxwell Allman, Itai Ashlagi, Amin Saberi, and Sophie H. Yu. 2025. From Signaling to Interviews in Random Matching Markets. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing* (Prague, Czechia) *(STOC '25)*. Association for Computing Machinery, New York, NY, USA, 1556–1567. https://doi.org/10.1145/3717823.3718158

Steffen Altmann, Anita Marie Glenny, Robert Mahlstedt, and Alexander Christopher Sebald. 2026. The Direct and Indirect Effects of Online Job Search Advice. *The Review of Economic Studies* (2026).

Joshua Angrist, Peter Hull, Parag A Pathak, and Christopher Walters. 2024a. Credible school value-added with undersubscribed school lotteries. *Review of Economics and Statistics* 106, 1 (2024), 1–19.

Joshua Angrist, Peter Hull, Parag A Pathak, and Christopher R Walters. 2024b. Race and the mismeasure of school quality. *American Economic Review: Insights* 6, 1 (2024), 20–37.

Dmitry Arkhangelsky and Wisse Rutgers. 2025. Evaluating Local Policies in Centralized Markets. https://doi.org/10.48550/arXiv.2510.20032 arXiv:2510.20032 [econ].

Nick Arnosti, Ramesh Johari, and Yash Kanoria. 2021. Managing congestion in matching markets. *Manufacturing & Service Operations Management* 23, 3 (2021), 620–636.

Dani Arribas-Bel, Seraphim Alvanides, Michael Batty, Andrew Crooks, Linda See, and Levi Wolf. 2021. Urban data/code: A new EP-B section. *Environment and Planning B: Urban Analytics and City Science* 48, 9 (Nov. 2021), 2517–2519. https://doi.org/10.1177/23998083211059670

Felipe Arteaga, Gregory Elacqua, Thomas Krussig, Carolina Méndez, and Christopher Neilson. 2022a. Can Information on School Attributes and Placement Probabilities Direct Search and Choice? Evidence from Choice Platforms in Ecuador and Peru. https://doi.org/10.18235/0004672

Felipe Arteaga, Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman. 2022b. Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice*. *The Quarterly Journal of Economics* 137, 3 (03 2022), 1791–1848. https://doi.org/10.1093/qje/qjac013 arXiv:https://academic.oup.com/qje/article-pdf/137/3/1791/44839848/qjac013.pdf

Itai Ashlagi, Mark Braverman, Yash Kanoria, and Peng Shi. 2020. Clearing matching markets efficiently: informative signals and match recommendations. *Management Science* 66, 5, 2163–2193.

Itai Ashlagi, Jiale Chen, Mohammad Roghani, and Amin Saberi. 2025. Stable matching with interviews. *arXiv preprint arXiv:2501.12503* (2025).

Itai Ashlagi, Anilesh K Krishnaswamy, Rahul Makhijani, Daniela Saban, and Kirankumar Shiragur. 2022. Assortment planning for two-sided sequential matching markets. *Operations Research* 70, 5 (2022), 2784–2803.

Amel Awadelkarim, Arjun Seshadri, Itai Ashlagi, Irene Lo, and Johan Ugander. 2023. Rank-heterogeneous preference models for school choice. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 47–56.

Mohammad Javad Azizi, Phebe Vayanos, Bryan Wilder, Eric Rice, and Milind Tambe. 2018. Designing fair, efficient, and interpretable policies for prioritizing homeless youth for housing resources. In *International conference on the integration of constraint programming, artificial intelligence, and operations research*. Springer, 35–51.

Sidhika Balachandar, Shuvom Sadhuka, Bonnie Berger, Emma Pierson, and Nikhil Garg. 2026. Urban incident prediction with graph neural networks: Integrating government ratings and crowdsourced reports. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40. 38194–38204.

Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrew Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. Improving refugee integration through data-driven algorithmic assignment. In *Science*, Vol. 359. 325–329.

Vince Bartle, Nicola Dell, and Nikhil Garg. 2025a. 'Shopping Around': An Experiment in Preferences and Incentives for Placing Long-term Patients. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–31.

Vince Bartle, Ashley Shearer, Alexandra Wroe, Nicola Dell, and Nikhil Garg. 2025b. Faster information for effective long-term discharge: A field study in adult foster care. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–29.

Hamsa Bastani, Kimon Drakopoulos, Vishal Gupta, Ioannis Vlachogiannis, Christos Hadjichristodoulou, Pagona Lagiou, Gkikas Magiorkinis, Dimitrios Paraskevis, and Sotirios Tsiodras. 2021. Efficient and targeted COVID-19 border testing via reinforcement learning. *Nature* 599, 7883 (2021), 108–113.

Luc Behaghel, Sofia Dromundo, Marc Gurgand, Yagan Hazard, and Thomas Zuber. 2024. *The potential of recommender systems for directing job search: A large scale experiment.* JSTOR.

Dimitris Bertsimas, Arthur Delarue, and Sebastien Martin. 2019. Optimizing schools' start time and bus routes. *Proceedings of the National Academy of Sciences* 116, 13 (2019), 5943–5948.

Omar Besbes, Yash Kanoria, and Akshit Kumar. 2025. Impact of Rankings and Personalized Recommendations in Marketplaces. *arXiv preprint arXiv:2506.03369* (2025).

Guillaume Bied, Solal Nathan, Elia Perennes, Morgane Hoffmann, Philippe Caillou, Bruno Crépon, Christophe Gaillac, and Michèle Sebag. 2023. Toward job recommendation for all. In *IJCAI 2023-The 32nd International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 5906–5914.

Guillaume Bied, Elia Perennes, Victor Alfonso Naya, Philippe Caillou, Bruno Crépon, Christophe Gaillac, and Michele Sebag. 2021. Congestion-avoiding job recommendation with optimal transport. In *FEAST workshop ECML-PKDD 2021*.

Thomas Blake and Dominic Coey. 2014. Why Marketplace Experimentation Is Harder Than It Seems: The Role of Test-Control Interference. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation (EC '14)*. ACM, 567–582. https://doi.org/10.1145/2600057.2602837

Zachary Bleemer and Basit Zafar. 2018. Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics* 157 (2018), 184–211.

Matteo Bobba, Veronica Frisancho, and Marco Pariguana. 2023. Perceived ability and school choices: Experimental evidence and scale-up effects. (2023).

Rishi Bommasani, Sarah H Bana, Kathleen A Creel, Dan Jurafsky, and Percy Liang. 2026. Algorithmic Monocultures in Hiring. *arXiv preprint arXiv:2605.27371* (2026).

Rishi Bommasani, Kathleen A Creel, Ananya Kumar, Dan Jurafsky, and Percy S Liang. 2022. Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? *Advances in neural information processing systems* 35 (2022), 3663–3678.

Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A Killian. 2021. Envisioning communities: a participatory approach towards AI for social good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 425–436.

Justin Boutilier, Jonas Oddur Jonasson, Hannah Li, and Erez Yoeli. 2025. Operational Dosage: Implications of Capacity Constraints for the Design and Interpretation of Experiments. https://doi.org/10.48550/arXiv.2407.21322 arXiv:2407.21322 [stat].

Bliss Broyard. 2023. NYC high school admissions rewards parents like me. https://ny.chalkbeat.org/2023/1/30/23574201/nyc-high-school-admissions-inequity-ethics Section: Front Page.

Julia Burdick-Will, Leela Gebo, and Alexandra D Williams. 2024. Anywhere but here: Neighborhood violence and local school preferences in Baltimore City. *Sociology of Education* 97, 1 (2024), 1–20.

Christopher Campos. 2024. *Social Interactions, Information, and Preferences for Schools: Experimental Evidence from Los Angeles*. Working Paper 33010. National Bureau of Economic Research. https://doi.org/10.3386/w33010

Serina Chang, Mandy L Wilson, Bryan Lewis, Zakaria Mehrab, Komal K Dudakiya, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, et al. 2021. Supporting covid-19 policy response with large-scale mobility-based modeling. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2632–2642.

Kuan-Ming Chen, Yu-Wei Hsieh, and Ming-Jen Lin. 2023. Reducing recommendation inequality via two-sided matching: A field experiment of online dating. *International Economic Review* 64, 3 (2023), 1201–1221.

Qinyi Chen, Jason Cheuk Nam Liang, Negin Golrezaei, and Djallel Bouneffouf. 2024. Interpolating Item and User Fairness in Multi-Sided Recommendations. In *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (Eds.), Vol. 37. Curran Associates, Inc., 50189–50229. https://doi.org/10.52202/079017-1589

Yilun Chen, Yash Kanoria, Akshit Kumar, and Wenxin Zhang. 2025. Feature-based dynamic matching. *Operations Research* (2025).

Erica Chiang, Katelyn Liu, Anum Ahmad, Zhi Liu, Gabriel Agostini, Tyler Gibson, Uma Bhandaram, and Nikhil Garg. 2026. The Forest from the Trees: Lessons on Public-Sector Deployment from an Academic Collaboration with NYC Department of Parks and Recreation. (2026).

Erica Chiang, Divya Shanmugam, Ashley N Beecy, Gabriel Sayer, Deborah Estrin, Nikhil Garg, and Emma Pierson. 2025. Learning disease progression models that capture health disparities. *Proceedings of the Conference on Health, Inference, and Learning* (2025).

Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on fairness, accountability and transparency*. PMLR, 134–148.

Konstantina Christakopoulou, Jaya Kawale, and Arindam Banerjee. 2017. Recommendation with Capacity Constraints. In *CIKM 2017 – Proceedings of the 2017 ACM Conference on Information and Knowledge Management*. ACM, 1439–1448. https://doi.org/10.1145/3132847.3133034

Sarah R Cohodes, Sean P Corcoran, Jennifer L Jennings, and Carolyn Sattin-Bajaj. 2025. When do informational interventions work? Experimental evidence from New York City high school choice. *Educational Evaluation and Policy Analysis* 47, 1 (2025), 208–236.

Barbara F Condliffe, Melody L Boyd, and Stefanie DeLuca. 2015. Stuck in school: How social context shapes school choice for inner-city students. *Teachers College Record* 117, 3 (2015), 1–36.

Sean P Corcoran, Jennifer L Jennings, Sarah R Cohodes, and Carolyn Sattin-Bajaj. 2018. *Leveling the Playing Field for High School Choice: Results from a Field Experiment of Informational Interventions*. Working Paper 24471. National Bureau of Economic Research. https://doi.org/10.3386/w24471

Viola Corradini. 2024. Information and Access in School Choice Systems: Evidence from New York City. (2024). https://opportunityinsights.org/wp-content/uploads/2024/11/JMP_Corradini.pdf

Viola Corradini and Clemence Idoux. 2023. *What matters in families' high school choices? Results from a high school post-application survey of parents*. Technical Report. New York City Department of Education. In collaboration with the NYCDOE Enrollment Research and Policy office, Carlos Flores, Brielle McDaniel, and Lianna Wright.

V. Corradini and C. Idoux. 2025a. School choice under uncertainty: search cost and disappointment aversion. (2025). Working paper.

Viola Corradini and Clemence M Idoux. 2025b. *Overcoming Racial Gaps in School Preferences: The Effect of Peer Diversity on School Choice*. Technical Report.

Kathleen Creel and Deborah Hellman. 2022. The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems. *Canadian Journal of Philosophy* 52, 1 (2022), 26–43.

Andrew Daw, Chloe Pache, and Angela Zhou. 2026. Due Process on Hold: A Queueing Framework for Improving Access in SNAP. *arXiv preprint arXiv:2605.15165* (2026).

Arthur Delarue, Zhen Lian, and Sebastien Martin. 2026. Algorithmic precision and human decision: A study of interactive optimization for school schedules. *Management Science* 72, 1 (2026), 148–166.

Evan Dong, Nikhil Garg, and Sarah Dean. 2026. Capacity Constraints Make Admissions Processes Less Predictable. *Proceedings of the AAAI Conference on Artificial Intelligence* 40, 45 (March 2026), 38385–38394. https://doi.org/10.1609/aaai.v40i45.41179

Umut Dur, Scott Duke Kominers, Parag A Pathak, and Tayfun Sönmez. 2018. Reserve design: Unintended consequences and the demise of Boston's walk zones. *Journal of Political Economy* 126, 6 (2018), 2457–2479.

Susan M Dynarski and Judith E Scott-Clayton. 2006. The cost of complexity in federal student aid: Lessons from optimal tax theory and behavioral economics. *National Tax Journal* 59, 2 (2006), 319–356.

Dean Eckles, Brian Karrer, and Johan Ugander. 2016. Design and Analysis of Experiments in Networks: Reducing Bias from Interference. *Journal of Causal Inference* 5, 1 (2016), 1–28. https://doi.org/10.1515/jci-2015-0021

Gregory Elacqua, Macarena Kutscher, Danielle Nascimento, Isabella Dias, and Juan Francisco Margitic. 2025. The Promise and Limits of Digital Nudges: Personalized School Recommendations in Recife's Centralized Admission Platform. (2025).

Juan F. Escobar and Alfonso Montes. 2025. Congestion and Information Design in Matching Markets. (2025).

Vivek F. Farias, Hao Li, Tianyi Peng, Xinyuyang Ren, Huawei Zhang, and Andrew Zheng. 2023. Correcting for Interference in Experiments: A Case Study at Douyin. https://doi.org/10.48550/arXiv.2305.02542 arXiv:2305.02542 [stat].

Luciano Floridi, Josh Cowls, Thomas C King, and Mariarosaria Taddeo. 2021. How to design AI for social good: Seven essential factors. In *Ethics, governance, and policies in artificial intelligence*. Springer, 125–151.

Matt Franchi, Nikhil Garg, Wendy Ju, and Emma Pierson. 2025. Bayesian modeling of zero-shot classifications for urban flood detection. *arXiv preprint arXiv:2503.14754* (2025).

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

Nikhil Garg. 2025. Heterogeneous participation and allocation skews: when is choice "worth it"? *Sigecom Exchanges* (2025).

Jennah Gosciak, Luke Boyce, Angelina Wang, and Allison Koenecke. 2026a. Scrutinizing Index-Based Risk Assessments: A Case Study in NYC Decision-making for Heat Emergency Management. *arXiv preprint arXiv:2605.17697* (2026).

Jennah Gosciak, Eric Giannella, Zhaowen Guo, Michael Chen, and Allison Koenecke. 2026b. LLMs in social services: How does chatbot accuracy affect human accuracy? *arXiv preprint arXiv:2603.11213* (2026).

Sophie Greenwood, Sudalakshmee Chiniah, and Nikhil Garg. 2024. User-item fairness tradeoffs in recommendations. *Advances in Neural Information Processing Systems* 37 (2024), 114236–114288.

Chantal A Hailey. 2022. Racial preferences for schools: Evidence from an experiment with White, Black, Latinx, and Asian parents and students. *Sociology of Education* 95, 2 (2022), 110–132.

Douglas N Harris and Matthew Larsen. 2015. What schools do families want (and why). *Education Research Alliance for New Orleans* 66 (2015).

Justine Hastings, Christopher A Neilson, and Seth D Zimmerman. 2015. *The effects of earnings disclosure on college enrollment decisions.* Technical Report. National Bureau of Economic Research.

Justine S Hastings, Christopher A Neilson, Anely Ramirez, and Seth D Zimmerman. 2016. (Un) informed college and major choice: Evidence from linked survey and administrative data. *Economics of Education Review* 51 (2016), 136–151.

Justine S Hastings, Richard Van Weelden, and Jeffrey Weinstein. 2007. *Preferences, Information, and Parental Choice Behavior in Public School Choice.* Working Paper 12995. National Bureau of Economic Research. https://doi.org/10.3386/w12995

Justine S Hastings and Jeffrey M Weinstein. 2008. Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics* 123, 4 (2008), 1373–1414.

YingHua He and Thierry Magnac. 2022. Application Costs and Congestion in Matching Markets. *The Economic Journal* 132, 648 (2022), 2918–2950. https://doi.org/10.1093/ej/ueac038

David Holtz, Felipe Lobel, Ruben Lobel, Inessa Liskovich, and Sinan Aral. 2025. Reducing interference bias in online marketplace experiments using cluster randomization: Evidence from a pricing meta-experiment on airbnb. *Management Science* 71, 1 (2025), 390–406.

Omar El Housni, Ulysse Hennebelle, and Alfredo Torrico. 2024. Two-sided assortment optimization: Adaptivity gaps and approximation algorithms. *arXiv preprint arXiv:2403.08929* (2024).

Caroline M Hoxby and Christopher Avery. 2012. *The missing" one-offs": The hidden supply of high-achieving, low income students.* Technical Report. National Bureau of Economic Research.

Caroline M. Hoxby and Sarah Turner. 2015. What High-Achieving Low-Income Students Know about College. *American Economic Review* 105, 5 (May 2015), 514–17. https://doi.org/10.1257/aer.p20151027

Ni Huang, Gordon Burtch, Yumei He, and Yili Hong. 2022. Managing congestion in a matching market via demand information disclosure. *Information Systems Research* 33, 4 (2022), 1196–1220.

Nicole Immorlica, Jacob Leshno, Irene Lo, and Brendan Lucier. 2020. Information acquisition in matching markets: The role of price discovery. *Available at SSRN 3705049* (2020).

Mobin Y Jeloudar, Irene Lo, and Katherine L Mentzer. 2025. Optimization Meets Participation: Iterative Zone Generation for School Assignment. In *Proceedings of the 5th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization.* 297–297.

Robert Jensen. 2010. The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics* 125, 2 (2010), 515–548.

Ashley Jochim, Michael DeArmond, Betheny Gross, and Robin Lake. 2014. How parents experience public school choice. *Center for Reinventing Public Education* (2014).

Ramesh Johari, Hannah Li, Inessa Liskovich, and Gabriel Y Weintraub. 2022. Experimental design in two-sided platforms: An analysis of bias. *Management Science* 68, 10 (2022), 7069–7089.

Ramesh Johari, Hannah Li, Anushka Murthy, and Gabriel Y. Weintraub. 2025a. When Does Interference Matter? Decision-Making in Platform Experiments. https://doi.org/10.48550/arXiv.2410.06580 arXiv:2410.06580 [stat].

Ramesh Johari, Orrie B. Page, and Gabriel Y. Weintraub. 2025b. Price Experimentation and Interference. https://doi.org/10.48550/arXiv.2310.17165 arXiv:2310.17165 [stat].

Ananya Joshi, Nolan Gormley, Richa Gadgil, Catalina Vajiac, Tina Townes, Roni Rosenfeld, and Bryan Wilder. 2026. Reducing Alert Fatigue Through AI Ranking: A Deployed Public Health Data Monitoring System. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40. 40006–40014.

Yash Kanoria and Daniela Saban. 2021. Facilitating the search for partners on matching platforms. *Management Science* 67, 10 (2021), 5990–6029.

Kleanthis Karakolios. 2025. Estimation of Treatment Effects in Matching Marketplaces Under Interference. (2025).

Mohit Karnani. 2024. *Essays on Empirical Matching Systems.* Massachusetts Institute of Technology.

Paula Kayongo, Glenn Sun, Jason Hartline, and Jessica Hullman. 2021. Visualization equilibrium. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 465–474.

Peter Kedron and Sarah Bardin. 2025. From Describing Problems to Solving Them: Communicating a Geographic Approach to Federal Evidence-Based Policymaking. *Annals of the American Association of Geographers* 0, 0 (Sept. 2025), 1–11. https://doi.org/10.1080/24694452.2025.2551856 _eprint: https://doi.org/10.1080/24694452.2025.2551856.

Elliot Kim, Avi Garg, Kenny Peng, and Nikhil Garg. 2025. Correlated errors in large language models. *arXiv preprint arXiv:2506.07962* (2025).

Jon Kleinberg and Manish Raghavan. 2021. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences* 118, 22 (2021), e2018340118.

Allison Koenecke, Eric Giannella, Robb Willer, and Sharad Goel. 2023. Popular support for balancing equity and efficiency in resource allocation: A case study in online advertising to increase welfare program awareness. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 494–506.

Meghana Kshirsagar, Caleb Robinson, Siyu Yang, Shahrzad Gholami, Ivan Klyuzhin, Sumit Mukherjee, Md Nasir, Anthony Ortiz, Felipe Oviedo, Darren Tanner, et al. 2021. Becoming good at ai for good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 664–673.

Tomás Larroucau, Ignacio Rios, Anaïs Fabre, and Christopher Neilson. 2024. College application mistakes and the design of information policies at scale. *Unpublished paper, Arizona State University, Tempe* (2024).

Hansol Lee, René F Kizilcec, and Thorsten Joachims. 2023. Evaluating a learned admission-prediction model as a replacement for standardized tests in college admissions. In *Proceedings of the tenth acm conference on learning@ scale*. 195–203.

Jinsook Lee, Emma Harvey, Joyce Zhou, Nikhil Garg, Thorsten Joachims, and René F Kizilcec. 2024. Algorithms for college admissions decision support: Impacts of policy change and inherent variability. *arXiv preprint arXiv:2407.11199* (2024).

Shuangning Li, Ramesh Johari, Xu Kuang, and Stefan Wager. 2024. Experimenting under Stochastic Congestion. https://doi.org/10.48550/arXiv.2302.12093 arXiv:2302.12093 [eess].

Zhi Liu, Uma Bhandaram, and Nikhil Garg. 2024a. Quantifying spatial under-reporting disparities in resident crowdsourcing. *Nature Computational Science* 4, 1 (2024), 57–65.

Zhi Liu and Nikhil Garg. 2024. Redesigning Service Level Agreements: Equity and Efficiency in City Government Operations. In *Proceedings of the 25th ACM Conference on Economics and Computation*. 309–309.

Zhi Liu, Sarah Rankin, and Nikhil Garg. 2024b. Identifying and Addressing Disparities in Public Libraries with Bayesian Latent Variable Modeling. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 20 (Mar. 2024), 22258–22265. https://doi.org/10.1609/aaai.v38i20.30231

Zhi Liu, Wenchang Zhu, Sarah Rankin, and Nikhil Garg. 2025. Optimizing library usage and browser experience: Application to the new york public library. *arXiv preprint arXiv:2503.23118* (2025).

Vahideh Manshadi, Scott Rodilitz, Daniela Saban, and Akshaya Suresh. 2025. Redesigning VolunteerMatch's Search Algorithm: Toward More Equitable Access to Volunteers. *Management Science* (2025).

Amelie Marian. 2023. Algorithmic Transparency and Accountability through Crowdsourcing: A Study of the NYC School Admission Lottery. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 434–443.

Yoosof Mashayekhi, Bo Kang, Jefrey Lijffijt, and Tijl De Bie. 2023. ReCon: Reducing congestion in job recommendation using optimal transport. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 696–701.

Ines Mergel, Helen Dickinson, Jari Stenvall, and Mila Gasco. 2024. Implementing AI in the public sector. *Public Management Review* (2024), 1–14.

Evan Munro. 2025. Causal Inference under Interference through Designed Markets. https://doi.org/10.48550/arXiv.2504.07217 arXiv:2504.07217 [econ].

Charlie Mydlarz, Praneeth Sai Venkat Challagonda, Bea Steers, Jeremy Rucker, Tega Brain, Brett Branco, Hannah E. Burnett, Amanpreet Kaur, Rebecca Fischman, Kathryn Graziano, Kendra Krueger, Elizabeth Hénaff, Véronëque Ignace, Erika Jozwiak, Jatin Palchuri, Polly Pierone, Paul Rothman, Ricardo Toledo-Crow, and Andrea I. Silverman. 2024. FloodNet: Low-Cost Ultrasonic Sensors for Real-Time Measurement of Hyperlocal, Street-Level Floods in New York City. *Water Resources Research* 60, 5 (2024), e2023WR036806. https://doi.org/10.1029/2023WR036806

Christopher A Neilson. 2026. A Decade of Centralized School Choice Admission in Chile: Achievements & Challenges. (2026).

NYC Public Schools. 2026a. Offer Chances. https://www.schools.nyc.gov/enrollment/enroll-grade-by-grade/high-school/offerchances. https://www.schools.nyc.gov/enrollment/enroll-grade-by-grade/high-school/offerchances Accessed: 2026-02-02.

NYC Public Schools. 2026b. *Specialized High Schools*. NYC Public Schools. https://www.schools.nyc.gov/enrollment/enroll-grade-by-grade/specialized-high-schools Accessed: 2026-02-09.

James G Oxley. 2006. *Matroid theory*. Vol. 3. Oxford University Press, USA.

Andrei Paleyes, Raoul-Gabriel Urma, and Neil D Lawrence. 2022. Challenges in deploying machine learning: a survey of case studies. *ACM computing surveys* 55, 6 (2022), 1–29.

Parag A Pathak and Peng Shi. 2021. How well do structural demand models work? Counterfactual predictions in school choice. *Journal of Econometrics* 222, 1 (2021), 161–195.

Kenny Peng and Nikhil Garg. 2024a. Monoculture in Matching Markets. In *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (Eds.), Vol. 37. Curran Associates, Inc., 81959–81991. https://doi.org/10.52202/079017-2605

Kenny Peng and Nikhil Garg. 2024b. Wisdom and Foolishness of Noisy Matching Markets. In *Proceedings of the 25th ACM Conference on Economics and Computation*. 675–675.

Kenny Peng, Emily Ryu, Jon Kleinberg, Eva Tardos, and Nikhil Garg. 2026. Connecting application behavior to undermatching in New York City school choice. *Nature Cities* (2026), 1–10.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. 2020. Performative prediction. In *International conference on machine learning*. PMLR, 7599–7609.

Andrew Perrault, Fei Fang, Arunesh Sinha, and Milind Tambe. 2020. Artificial intelligence for social impact: Learning and planning in the data-to-deployment pipeline. *Ai Magazine* 41, 4 (2020), 3–16.

Canice Prendergast. 2017. How food banks use markets to feed the poor. *Journal of Economic Perspectives* 31, 4 (2017), 145–162.

Naveen Janaki Raman, Jingwu Tang, Zhiyu Chen, Zheyuan Ryan Shi, Sean Hudson, Ameesh Kapoor, and Fei Fang. 2026. RescueLens: LLM-Powered Triage and Action on Volunteer Feedback for Food Rescue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40. 40092–40100.

Ignacio Rios, Daniela Saban, and Fanyin Zheng. 2023. Improving match rates in dating markets through assortment optimization. *Manufacturing & Service Operations Management* 25, 4 (2023), 1304–1323.

Ignacio Rios and Alfredo Torrico. 2023. Platform Design in Curated Dating Markets. *arXiv preprint arXiv:2308.02584* (2023).

Carolyn Sattin-Bajaj, Jennifer L Jennings, Sean P Corcoran, Elizabeth Christine Baker-Smith, and Chantal Hailey. 2018. Surviving at the street level: How counselors' implementation of school choice policy shapes students' high school destinations. *Sociology of Education* 91, 1 (2018), 46–71.

Devansh Saxena, Ji-Youn Jung, Jodi Forlizzi, Kenneth Holstein, and John Zimmerman. 2025. AI mismatches: identifying potential algorithmic harms before AI development. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.

Alexander Schrijver et al. 2003. *Combinatorial optimization: polyhedra and efficiency*. Vol. 24. Springer.

Jonathan Schroeder, David Van Riper, Steven Manson, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2025. IPUMS National Historical Geographic Information System: Version 20.0. [dataset]. https://doi.org/10.18128/D050.V20.0

Friso Selten and Bram Klievink. 2024. Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly* 41, 1 (2024), 101885.

Peng Shi. 2025. Optimal match recommendations in two-sided marketplaces with endogenous prices. *Management Science* 71, 9 (2025), 7431–7448.

Zheyuan Ryan Shi, Leah Lizarondo, and Fei Fang. 2021. A recommender system for crowdsourcing food rescue platforms. In *Proceedings of the web conference 2021*. 857–865.

Jeremy Singer. 2025. School Choice, Socioeconomic Status, and Stratified Enrollment in Urban Districts: Evidence From Detroit. *Educational Evaluation and Policy Analysis* 47, 3 (2025), 821–846. https://doi.org/10.3102/01623737241254783

Yi Su, Magd Bayoumi, and Thorsten Joachims. 2022. Optimizing rankings for recommendation in matching markets. In *Proceedings of the ACM Web Conference 2022*. 328–338.

Sabina Tomkins, Joshua Grossman, Lindsay Page, and Sharad Goel. 2023. Showing high-achieving college applicants past admissions outcomes increases undermatching. *Proceedings of the National Academy of Sciences* 120, 45 (2023), e2306017120.

Alfredo Torrico, Margarida Carvalho, and Andrea Lodi. 2020. Multi-agent assortment optimization in sequential matching markets. *arXiv preprint arXiv:2006.04313* (2020).

Jon Valant and Lindsay H. Weixler. 2022. Informing School-Choosing Families About Their Options: A Field Experiment From New Orleans. *Educational Evaluation and Policy Analysis* 44, 4 (2022), 608–637. https://doi.org/10.3102/01623737221086305

Christopher R Walters. 2018. The demand for effective charter schools. *Journal of Political Economy* 126, 6 (2018), 2179–2223.

Irene Wapnir, Itai Ashlagi, Alvin E Roth, Erling Skancke, Akhil Vohra, Irene Lo, and Marc L Melcher. 2021. Explaining a potential interview match for graduate medical education. *Journal of Graduate Medical Education* 13, 6 (2021), 764–767.

Bryan Wilder and Angela Zhou. 2025. Fostering the Ecosystem of AI for Social Impact Requires Expanding and Strengthening Evaluation Standards. *arXiv preprint arXiv:2510.18238* (2025).

Matthew Wiswall and Basit Zafar. 2015. Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies* 82, 2 (2015), 791–824.

Xiaoyang Ye. 2024. Improving college choice in centralized admissions: Experimental evidence on the importance of precise predictions. *Education Finance and Policy* 19, 2 (2024), 308–340.

Dongping Zhang, Jason Hartline, and Jessica Hullman. 2024. Designing shared information displays for agents of varying strategic sophistication. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–34.

Eman Zied Abozied, Caitlin Robinson, Rachel Franklin, Kate Court, and Jack Roberts. 2025. A spatial decision support framework for equitable sensor network distribution in the smart city. *The Geographical Journal* 191, 1 (2025), e12605. https://doi.org/10.1111/geoj.12605

Anneke Zuiderwijk, Yu-Che Chen, and Fadi Salem. 2021. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government information quarterly* 38, 3 (2021), 101577.

A ADDITIONAL RELATED WORK

Algorithmic tools to support the delivery of social and government services while accounting for disparities. Recent work applies, and examines the performance of, predictive modeling and optimization methods in a range of domains, especially in New York City. These include urban planning and incident prediction [Agostini et al., 2024, Balachandar et al., 2026, Franchi et al., 2025, Gosciak et al., 2026a, Liu et al., 2024a, Liu and Garg, 2024, Mydlarz et al., 2024, Zied Abozied et al., 2025]; allocation of housing and social services [Azizi et al., 2018, Chouldechova et al., 2018, Daw et al., 2026, Gosciak et al., 2026b, Koenecke et al., 2023], library resources [Liu et al., 2024b, 2025], and food [Prendergast, 2017, Raman et al., 2026, Shi et al., 2021]; healthcare [Adams et al., 2022, Bartle et al., 2025a,b, Bastani et al., 2021, Chang et al., 2021, Chiang et al., 2025, Joshi et al., 2026]; education [Bertsimas et al., 2019, Delarue et al., 2026, Lee et al., 2023, 2024, Tomkins et al., 2023]; and refugee resettlement [Ahani et al., 2024, Bansak et al., 2018]. One consistent concern in such work—including in our setting—is how administrative burdens and resulting heterogeneous participation may have equity implications in sociotechnical systems [Garg, 2025].

A line of work builds on these developments to explore the actual process of translating research into real-world impact. These papers examine challenges in deployments [Kedron and Bardin, 2025, Paleyes et al., 2022, Saxena et al., 2025, Selten and Klievink, 2024], propose principles and strategies for more effective and responsible implementation in practice [Bondi et al., 2021, Chiang et al., 2026, Floridi et al., 2021, Kshirsagar et al., 2021, Perrault et al., 2020], and argue broadly for more research on implementation-specific issues or different academic evaluation norms [Arribas-Bel et al., 2021, Mergel et al., 2024, Wilder and Zhou, 2025, Zuiderwijk et al., 2021]. Our work provides one case study on the deployment and evaluation of such an algorithmic system.

Experimentation under interference. A growing literature also studies experimentation under interference in markets with scarce resources and develops methods for *computing treatment effects* in these settings [Arkhangelsky and Rutgers, 2025, Blake and Coey, 2014, Boutilier et al., 2025, Eckles et al., 2016, Farias et al., 2023, Holtz et al., 2025, Johari et al., 2022, 2025a,b, Karakolios, 2025, Karnani, 2024, Li et al., 2024, Munro, 2025]. For example, if the intervention increases demand for a program among the treatment group, then acceptance rates at that program may decrease for the control group, inflating a naively estimated treatment effect on match outcomes. Our work is complementary to this literature, which focuses on randomization schemes and debiasing estimators; we focus on the intervention design itself to *minimize interference* by preserving offer likelihoods for applicants with ex ante *High* likelihoods.

Our ability to simulate match outcomes given application behavior allows us to characterize interference effects and optimize our intervention accordingly. Using this capability, we further run scale-up analyses to explore how large our intervention could be while still bounding interference, analogous to analyses in Al-Ubaydli et al. [2017], Allende et al. [2019], Larroucau et al. [2024].

B THEORETICAL PROOFS

B.1 Coverage and proof of Theorem 3.2

We first define *Coverage*, which captures how nudges are distributed.

Definition 7 (Coverage). Coverage at k under X is the number of students who receive at least k nudges.

$$\mathrm{Cov}_{\mathrm{k}}(X) = \sum_{i \in \mathcal{S}} \mathbf{1}\{k_i(X) \geq k\}.$$

For example, $\mathrm{Cov}_1(X)$ is the number of students who receive at least one nudge. We now define lexicographic maximum coverage, a property of an allocation that characterizes first maximizing the number of students who receive at least one nudge, then the number receiving two, and so on.

Definition 8 (Lexicographic maximum coverage). An allocation X^* has *lexicographic maximum coverage* among all feasible allocations given Z, C if there does not exist any feasible $X$and any k such that $\mathrm{Cov_k}(X) > \mathrm{Cov_k}(X^*)$ and $\mathrm{Cov_{k'}}(X) = \mathrm{Cov_{k'}}(X^*)$ for all $k' < k$.

Definition 9 (Prefix-maximal coverage). An allocation X^* has *prefix-maximal coverage* among all feasible allocations given Z, C if, for every feasible allocation X and every $k \geq 1$,

$$\sum_{r=1}^{k} \mathrm{Cov_r}(X^*) \geq \sum_{r=1}^{k} \mathrm{Cov_r}(X).$$

Note that prefix-maximal coverage implies lexicographic maximum coverage.

We show that, perhaps surprisingly, *any* strictly concave u yields prefix-maximal coverage, and therefore lexicographic maximum coverage.

Theorem B.1 (Concavity maximizes coverage). *Let u be increasing and strictly discretely concave. Then, X is an optimum of $\boldsymbol{CNA}(C, u) \iff$ it has prefix-maximal coverage.*

We first prove Theorem B.1 by showing that the feasible allocations induce a matroid, so the greedy algorithm (which produces prefix-maximal coverage) is optimal. For intuition, we also provide a direct construction in which any allocation that does not maximize Cov_1 can be improved by reallocating nudges from students with multiple nudges to those with no nudges, increasing the concave objective and Cov_1.[23] We then use the resulting coverage property to prove Theorem 3.2.

Setup. Suppose there are n students and m nudge-eligible programs. Recall the feasibility constraints

$$\sum_{i=1}^{n} X_{ij} \leq C_j \quad \forall j, \qquad X_{ij} \leq Z_{ij} \quad \forall i, j, \qquad X_{ij} \in \{0, 1\}.$$

Let $k_i(X) := \sum_j X_{ij}$, $\widehat{C}_j := \min\{C_j, \sum_i Z_{ij}\}$, and $B := \sum_{j=1}^{m} \widehat{C}_j$. Define discrete marginals $\Delta(t) := u(t) - u(t-1)$ for $t \geq 1$. Without loss of generality, assume $u(0) = 0$.

Theorem B.1 (Concavity maximizes coverage). *Let u be increasing and strictly discretely concave. Then, X is an optimum of $\boldsymbol{CNA}(C, u) \iff$ it has prefix-maximal coverage.*

Proof. We first construct a network flow problem in which (1) feasible integral flows encode exactly the feasible nudge allocations; and (2) the subsets of student nodes that can be simultaneously included in an integer flow form the independent sets of a matroid. Then, we use the fact that the greedy algorithm is optimal for maximizing linear weights over matroids to show that the optimizer of $\mathbf{CNA}(C, u)$ corresponds to a prefix-maximal nudge allocation when u is strictly concave.

Step 1: Build a network whose feasible flows encode feasible nudge allocations. The network is illustrated in Figure 6. We construct a layered graph with a source, student-copy nodes, intermediate pair nodes, program nodes, and a sink.

A feasible nudge allocation X assigns each student a set of nudges to distinct programs where they are nudge-eligible, subject to nudge capacities. To model this, we create multiple copies of each student, where each copy represents one potential nudge for that student. Pair nodes enforce

[23]Note that the result is trivially true if $Z_{ij} = 1$ for all i, j, since then we can arbitrarily transfer nudges to those without nudges; but for general Z we need to show transfer feasibility, which we do so via an alternating path construction.

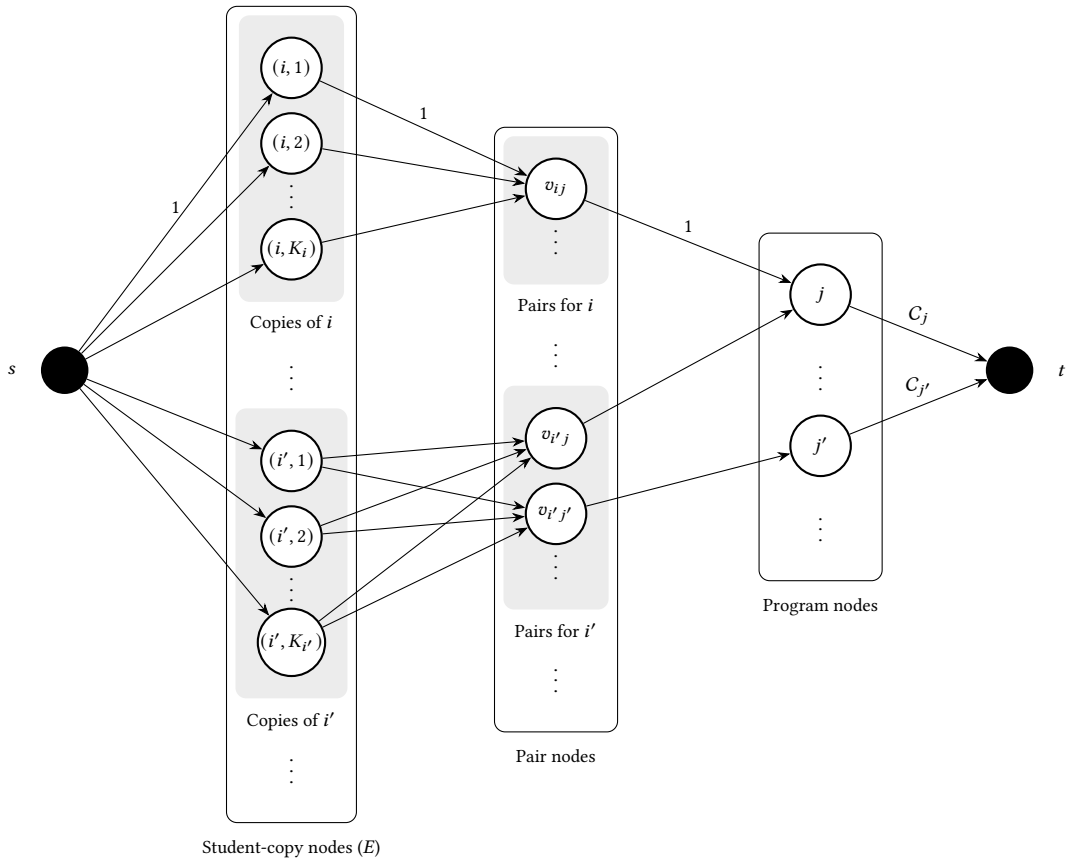


Fig. 6. **Diagram for proof of Theorem B.1: flow network encoding feasible nudge allocations.** Each unit of $s \to t$ flow corresponds to one nudge. The first column contains $K_i = \sum_j Z_{ij}$ copies of each student i, where each copy (i, r) represents a potential nudge. The second column contains pair nodes v_{ij} for eligible pairs (i, j) with $Z_{ij} = 1$. The third column contains one node for each program j, with an edge of capacity C_j to the sink. A feasible flow corresponds to a feasible nudge allocation: each student copy sends flow to at most one program, flow is allowed only between nudge-eligible pairs (i, j), each student sends at most 1 unit of flow to any program, and each program j receives at most C_j nudges. Edges $s \to (i, r)$ enforce "each copy used at most once." Edges $(i, r) \to v_{ij}$ correspond to eligible nudges Z_{ij}. Edges $v_{ij} \to j$ enforce that a student cannot be nudged to the same program twice (only one copy of student i can send flow to v_{ij}). Edges $j \to t$ enforce program nudge capacities.

that no student is nudged to the same program twice (at most one copy of student i can send flow to program j).

Formally, for each student i, define the maximum possible number of distinct nudges they could receive as

$$K_i := \sum_{j=1}^{m} Z_{ij}.$$

Create a *copy ground set*

$$E := \{(i, r) : i \in [n],\ r \in [K_i]\}.$$

A unit of flow through node (i, r) represents allocating one nudge to student i.

We build a directed graph $D = (V, A)$ with source s and sink t as follows:

- *Student-copy nodes*: for each copy element $e = (i, r) \in E$, create a node (i, r) and add an edge $s \to (i, r)$ with capacity 1.
- *Pair nodes*: for each eligible pair (i, j) with $Z_{ij} = 1$, create a node v_{ij}. For every copy (i, r) and program j with $Z_{ij} = 1$, add an edge $(i, r) \to v_{ij}$ with capacity 1.
- *Program nodes*: for each program j, create a node j. Add edges $v_{ij} \to j$ with capacity 1 for all i. Add an edge $j \to t$ with capacity C_j.

Pair nodes enforce that each student can be nudged to a given program at most once: since the edge $v_{ij} \to j$ has capacity 1, at most one copy of student i can send flow to program j. The edges $j \to t$ enforce program capacities C_j, and the edges $s \to (i, r)$ ensure that each copy of a student receives at most one nudge.

Feasible flows in this network correspond exactly to feasible nudge allocations X (see caption of Figure 6 for a full mapping of flow constraints to nudge allocation constraints).

Step 2: The matchable sets of student-copy nodes form a matroid. To state the matroid property cleanly, we convert the program capacities to unit-capacity terminals. For each program j, replace the $j \to t$ edge of capacity C_j with C_j distinct terminal vertices $t_{j,1}, \ldots, t_{j,C_j}$, and add unit-capacity edges $j \to t_{j,r}$ for $r \in [C_j]$. Let $T := \{t_{j,r} : j \in [m],\ r \in [C_j]\}$ denote the set of terminals.

For a subset $S \subseteq E$ of student-copy nodes, call S *linkable* if there exist $|S|$ edge-disjoint directed paths in D that start at the elements of S and end at distinct terminals in T. (Equivalently, there exists an integral flow that sends one unit from the source through each element of S and routes these units to distinct terminals in T.)

Define

$$\mathcal{I} \ := \ \{S \subseteq E : \ S \text{ is linkable to } T \text{ in } D\}.$$

A standard result in matroid theory asserts that $(E, \mathcal{I})$ is a matroid (a gammoid, specifically) [Schrijver et al., 2003]. Thus the family of student-copy sets that can simultaneously be matched to terminals (which correspond to nudges) forms a matroid on ground set E. (In the directed line graph of D, these paths correspond to vertex-disjoint paths from vertices representing edges $s \to e$ ($e \in S$) to vertices representing edges entering T).

For each program j, choose any $\widehat{C}_j$ eligible applicants. Because applicants have no nudge limit, these choices form a feasible allocation using all B nudges. Hence this matroid has rank B and allocations using all B nudges correspond exactly to bases of the matroid.

Step 3: Linearize the concave objective via marginal weights on copies. Assign each element $e = (i, r) \in E$ the weight

$$w(i, r) := \Delta(r) = u(r) - u(r - 1).$$

Normalization. We may assume $u(0) = 0$ without loss of generality: replacing u by $\tilde{u}(r) := u(r) - u(0)$ shifts the objective by the constant $n\,u(0)$ and does not change monotonicity or discrete concavity.

Canonical copies. Because all copies (i, r) of a fixed student i have identical outgoing edges in D, linkability of a set $S \subseteq E$ depends only on the multiplicities $k_i(S) := \big|\{r : (i, r) \in S\}\big|$. In particular, if $S \in \mathcal{I}$ and S' is obtained by permuting which copies of student i are selected (holding k_i fixed for all i), then $S' \in \mathcal{I}$ as well.

Therefore, among optimal solutions we may restrict attention to *canonical* sets that, for each student i, select the first $k_i(S)$ copies $\{(i,1),\dots,(i,k_i(S))\}$. For such a canonical set,

$$\sum_{(i,r)\in S} w(i,r) = \sum_{i=1}^{n}\sum_{r=1}^{k_i(S)} \Delta(r) = \sum_{i=1}^{n} u(k_i(S)),$$

using the telescoping identity $u(k) = \sum_{r=1}^{k}(u(r)-u(r-1))$ and $u(0)=0$.

Step 4: Greedy optimality and prefix maximality. By the matroid greedy theorem (see, e.g., Oxley [2006], Schrijver et al. [2003]), a maximum-weight base is obtained by sorting elements by nonincreasing weight and adding each element if it preserves independence.

If u is strictly discretely concave, then $\Delta(1) > \Delta(2) > \cdots$, so weights strictly decrease in level index r.

Partition the ground set by level: $E^{(t)} := \{(i,t) : t \le K_i\}$. Since $w(i,t)$ depends only on t, the greedy algorithm processes all of $E^{(1)}$ first, then $E^{(2)}$, etc. Let $S^\star$ be any maximum-weight base. Under strictly decreasing weights, a standard consequence of the matroid greedy theorem is that for every $t \ge 1$, $S^\star$ maximizes the prefix intersection size

$$\left| S^\star \cap \bigcup_{r=1}^{t} E^{(r)} \right|$$

among all bases. Because we may restrict attention to canonical bases, and for any canonical base S,

$$\left| S \cap \bigcup_{r=1}^{t} E^{(r)} \right| = \sum_{r=1}^{t} \mathrm{Cov}_r(S),$$

every optimum $S^\star$ therefore satisfies, for every base S and every $t \ge 1$,

$$\sum_{r=1}^{t} \mathrm{Cov}_r(S^\star) \ge \sum_{r=1}^{t} \mathrm{Cov}_r(S).$$

Any feasible allocation can be extended to a base, which weakly increases each prefix sum, so this inequality also holds relative to every feasible allocation. (Intuitively: $\Delta(t) > \Delta(t+1) \ge \Delta(t+2) \ge \cdots$, and so the greedy algorithm first picks as many $t=1$ elements as possible, and only picks a $t=2$ element (a second nudge for a student) if it cannot pick any more $t=1$ elements, and so on.)

This prefix maximization implies lexicographic maximum coverage. A student contributes a copy to $S \cap E^{(r)}$ if and only if $k_i(S) \ge r$. Thus, for $t=1$, $\left|S^\star \cap E^{(1)}\right| = \mathrm{Cov}_1(S^\star)$ is maximized. Holding Cov_1 fixed, the above then maximizes $\left|S^\star \cap E^{(2)}\right| = \mathrm{Cov}_2(S^\star)$, the number of students with at least 2 nudges (2 copies with flow through them). Continuing, the above greedy solutions maximize the tail vector $(\mathrm{Cov}_1(S), \mathrm{Cov}_2(S), \dots)$ lexicographically, and are optimal by the matroid greedy theorem. Thus, every optimal solution has prefix-maximal (and therefore lexicographic maximum) coverage.

The reverse direction is straightforward: if S is not optimal, then there exists an optimal solution $S^\star$ with strictly higher weight. Let t be the smallest level at which S and $S^\star$ differ. Then, $S^\star$ has more elements in $E^{(t)}$ than S, so $\mathrm{Cov}_t(S^\star) > \mathrm{Cov}_t(S)$, contradicting prefix maximality of coverage. □

B.1.1 Intuition for Theorem B.1. The proof of Theorem B.1 uses matroid-based arguments. For intuition on the result, next we prove a weaker result (that with strict concavity, every optimum maximizes Cov_1) more directly, without invoking such machinery.

A nudge allocation is a binary matrix $X \in \{0,1\}^{n\times m}$, where $X_{ij} = 1$ indicates nudging student i to program j, subject to the program-level nudge capacities and eligibility constraints

$$\sum_{i=1}^{n} X_{ij} \le C_j \quad \forall j, \qquad X_{ij} \le Z_{ij} \quad \forall i, j. \tag{2}$$

Given X, define each student's nudge list size $k_i(X) = \sum_{j=1}^{m} X_{ij}$ and the coverage

$$\mathrm{Cov}(X) = \sum_{i=1}^{n} \mathbf{1}\{k_i(X) \ge 1\}.$$

Let $N(X) = \sum_{i,j} X_{ij} = \sum_i k_i(X)$ be the total number of nudges.

Definition 10 (Maximal feasible coverage under full utilization). For $C \le \bar{c}$, let $B = \sum_{j=1}^{m} C_j$. Define the maximal feasible coverage among full-utilization (equivalently, linear-optimal) allocations as

$$\mathrm{Cov}_{\max}(Z, C) \;:=\; \max\Big\{\mathrm{Cov}(X) : X \text{ satisfies (2) and } N(X) = B\Big\}.$$

As a baseline, one can maximize the total number of nudges (a linear objective):

$$\max_X \sum_{i=1}^{n} k_i(X) \quad \text{s.t. (2).} \tag{3}$$

Now instead fix an increasing, discretely concave $u : \mathbb{Z}_{\ge 0} \to \mathbb{R}_{\ge 0}$. Let $u(0) = 0$ without loss of generality. The concave surrogate objective solves

$$\max_X \sum_{i=1}^{n} u\big(k_i(X)\big) \quad \text{s.t. (2).} \tag{4}$$

Theorem B.2 (Concavity maximizes number of students with at least one nudge). *For $C \le \bar{c}$, fix any increasing, discretely concave u with $u(0) = 0$.*

(1) *There exists an optimizer $X^\star$ of* (4) *satisfying*

$$\mathrm{Cov}(X^\star) = \mathrm{Cov}_{\max}(Z, C).$$

In particular, for any optimizer $\tilde{X}$ of the linear problem (3)*, we have* $\mathrm{Cov}(X^\star) \ge \mathrm{Cov}(\tilde{X})$.

(2) *If, in addition, u has* strictly diminishing returns *between the first and second nudge, i.e.,*

$$u(1) - u(0) \;>\; u(2) - u(1),$$

then every *optimizer $X^\star$ of* (4) *satisfies* $\mathrm{Cov}(X^\star) = \mathrm{Cov}_{\max}(Z, C)$.

Proof. *High-level intuition.* We argue by contradiction. Suppose a full-utilization maximum-coverage allocation Y covers more students than a concave-optimal allocation X. Note that we can restrict attention to full-utilization concave-optimal allocations; $C \le \bar{c}$ implies $|\{i : Z_{ij} = 1\}| \ge C_j$ for every program j, so if some program j has unused capacity (i.e., $\sum_i X_{ij} < C_j$), then there exists an eligible student i with $X_{ij} = 0$; setting $X_{ij} = 1$ increases $k_i(X)$ by one and weakly increases $\sum_i u(k_i(X))$. Repeating this operation yields a full-utilization allocation with objective weakly larger.

Then, we show that, from the concave-optimal allocation, it is feasible to move a nudge from a student with at least 2 nudges to a student with 0 nudges, via an alternating-path swap. This results in increasing coverage and (by diminishing returns) not decreasing the concave objective (strictly increasing it under the strict condition).

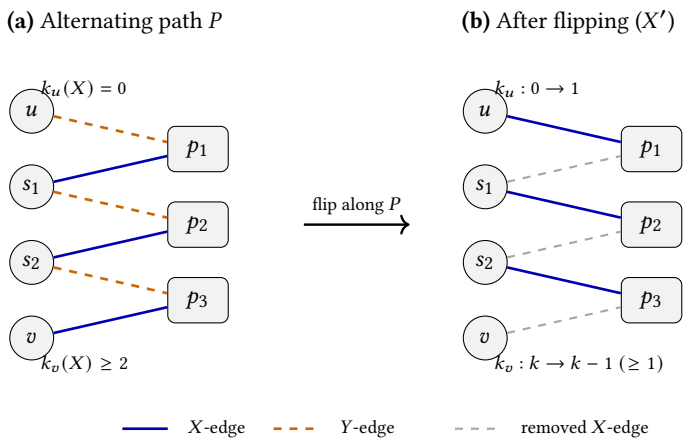


Fig. 7. **Alternating-path exchange in Theorem B.2** Panel (a) highlights an alternating path P in the symmetric difference $E(X) \triangle E(Y)$ between a concave-optimal allocation X and a full-utilization maximum-coverage allocation Y. Flipping X along P (panel (b)) preserves program caps, increases coverage by one at the start vertex u, and does not decrease the concave objective by discrete concavity.

Step 1 (Pick benchmark allocations). Let $\mathcal{X}^\star$ be the set of optimizers of the concave problem (4), and choose $X \in \mathcal{X}^\star$ that maximizes $\mathrm{Cov}(X)$ over $\mathcal{X}^\star$. By monotonicity of u and the fact that $C \leq \overline{c}$ makes full utilization feasible, we may assume X is full utilization, i.e., $N(X) = B$.

Let Y be the full-utilization allocation with $\mathrm{Cov}(Y) = \mathrm{Cov}_{\max}(Z, C)$ (such a Y exists by Definition 10). Assume for contradiction that $\mathrm{Cov}(Y) > \mathrm{Cov}(X)$.

Step 2 (Build the symmetric-difference graph). Let $E(X) = \{(i, j) : X_{ij} = 1\}$ and define the symmetric difference subgraph

$$H = (V,\ E(X) \triangle E(Y)),$$

where $\triangle$ denotes *symmetric difference*: for sets A, B, $A \triangle B = (A \setminus B) \cup (B \setminus A)$. Thus $E(X) \triangle E(Y)$ consists of the edges present in exactly one of X and Y (edges common to both are omitted). We color edges in $E(X) \setminus E(Y)$ as X and edges in $E(Y) \setminus E(X)$ as Y. Here V consists of all students and all programs.

Step 3 (Decompose H into alternating cycles/paths). Since both X and Y are full utilization, every program vertex j has equal X-degree and Y-degree in H. To see this, note that $\sum_i X_{ij} = \sum_i Y_{ij} = C_j$ and edges common to X and Y do not appear in H, so the number of X-only edges incident to j equals the number of Y-only edges incident to j. Equivalently, writing $\deg_H^{(X)}(v)$ and $\deg_H^{(Y)}(v)$ for the number of incident X-colored and Y-colored edges at a vertex v in H, we have $\deg_H^{(X)}(j) = \deg_H^{(Y)}(j)$ for every program j.

This balance implies that H can be decomposed into edge-disjoint alternating cycles and alternating paths. Here an *alternating* cycle/path is one whose edges alternate in color $(Y, X, Y, X, \dots)$, and *edge-disjoint* means that the components partition $E(H)$. To construct such a decomposition, fix an arbitrary pairing at each program vertex j that matches each incident Y-edge to a distinct incident X-edge (possible since $\deg_H^{(Y)}(j) = \deg_H^{(X)}(j)$). Starting from any unused edge, follow a maximal alternating walk that (i) at programs always continues along the paired opposite-color edge and (ii) at students takes any unused opposite-color edge if one exists; this walk cannot get

stuck at a program and therefore yields either an alternating cycle or an alternating path with student endpoints. Delete its edges and repeat until no edges remain.

We will orient each alternating path to start with a Y-edge (equivalently, to end with an X-edge); since H is bipartite, every student-to-student path has even length and hence this orientation is well-defined. If the end of one oriented path equals the start of another, concatenate them and repeat; then, no vertex is simultaneously the end of one path and the start of another. In this resulting decomposition, any end vertex v of an oriented alternating path satisfies $\deg_H^{(X)}(v) > \deg_H^{(Y)}(v)$: cycles and internal vertices contribute equal numbers of incident X- and Y-edges, while an oriented path contributes one extra incident Y-edge at its start and one extra incident X-edge at its end. See Figure 7 for an illustration.

Step 4 (Locate a path from a student with 0 nudges in x to a student with at least 2 nudges in X). Recall that $k_i(X) = \sum_{j=1}^{m} X_{ij}$ denotes student i's nudge list size under allocation X. Let $S_X = \{i : k_i(X) > 0\}$ and $S_Y = \{i : k_i(Y) > 0\}$ denote the sets of covered students under X and Y, so $\mathrm{Cov}(X) = |S_X|$ and $\mathrm{Cov}(Y) = |S_Y|$. Since $\mathrm{Cov}(Y) > \mathrm{Cov}(X)$, we have $|S_Y \setminus S_X| > |S_X \setminus S_Y|$.

Define the "newly covered" and "fragile uncovered" sets

$$U = S_Y \setminus S_X = \{i : k_i(X) = 0,\ k_i(Y) > 0\}, \qquad L = \{i : k_i(X) = 1,\ k_i(Y) = 0\}.$$

Here U are students covered by Y but not by X, while $L \subseteq S_X \setminus S_Y$ are students who have exactly one X-edge and no Y-edge (so removing a single X-edge would make them uncovered). Since $|U| = |S_Y \setminus S_X| > |S_X \setminus S_Y| \geq |L|$, we have $|U| > |L|$.

Consider the edge-disjoint decomposition of H into alternating cycles and alternating paths from Step 3, with every alternating path oriented to start with a Y-edge (and hence to end with an X-edge). For a student i, edges common to X and Y do not appear in H, so

$$k_i(X) - k_i(Y) = \deg_H^{(X)}(i) - \deg_H^{(Y)}(i).$$

In particular, each $i \in U$ has $\deg_H^{(Y)}(i) = k_i(Y) \geq 1$ and $\deg_H^{(X)}(i) = 0$, so every $i \in U$ must be the start of at least one alternating path. Each $i \in L$ has $\deg_H^{(X)}(i) = 1$ and $\deg_H^{(Y)}(i) = 0$, so i is the end of exactly one alternating path (since we need a unique incoming edge in X for an alternating path to end at i). Since $|U| > |L|$, there exists an alternating path P whose start vertex $u \in U$ and whose end vertex v is not in L. By the endpoint property from Step 3, we have $\deg_H^{(X)}(v) > \deg_H^{(Y)}(v)$, and hence $k_v(X) > k_v(Y)$ by the displayed identity. Finally, since $v \notin L$, this implies $k_v(X) \geq 2$ (otherwise $k_v(X) = 1$ forces $k_v(Y) = 0$ and hence $v \in L$).

Step 5 (Flip along the path). We have located an alternating path P in H starting at $u \in U$ with $k_u(X) = 0$ and ending at v with $k_v(X) \geq 2$. We can now reallocate nudges along this path to increase coverage without decreasing the concave objective (and strictly increasing it under the additional strict-diminishing-returns condition).

Define X' by flipping X along P: set $X'_{ij} = 1$ for edges $(i, j) \in E(Y) \cap P$, set $X'_{ij} = 0$ for edges $(i, j) \in E(X) \cap P$, and keep all other entries the same. By alternation, every program on P loses exactly one incident X-edge and gains exactly one incident Y-edge, so X' remains feasible under program capacities. Eligibility is also preserved since every added edge lies in $E(Y)$ and Y satisfies $Y_{ij} \leq Z_{ij}$. All internal students on P lose and gain one nudge, so their list sizes are unchanged; the endpoints satisfy

$$k_u(X') = k_u(X) + 1 = 1, \qquad k_v(X') = k_v(X) - 1 \geq 1.$$

Thus $\mathrm{Cov}(X') = \mathrm{Cov}(X) + 1$.

Finally, only k_u and k_v change, so the concave objective changes by

$$\big(u(1) - u(0)\big) - \big(u(k_v(X)) - u(k_v(X) - 1)\big) \ \geq\ 0$$

by discrete concavity, with strict inequality under the additional condition $u(1)-u(0) > u(2)-u(1)$ since $k_v(X) \geq 2$.

We now close the contradiction. Since X is concave-optimal and the objective under X' is weakly larger, X' is also concave-optimal. But $\mathrm{Cov}(X') > \mathrm{Cov}(X)$ contradicts the choice of X as the maximum-coverage optimizer in $\mathcal{X}^\star$. Therefore $\mathrm{Cov}(X) = \mathrm{Cov}_{\max}(Z, C)$, proving part (1).

For part (2), assume $u(1)-u(0) > u(2)-u(1)$. Then discrete concavity implies $u(k)-u(k-1) \leq u(2)-u(1)$ for all $k \geq 2$, so the objective change above is strictly positive since $k_v(X) \geq 2$. Thus no optimizer of (4) can satisfy $\mathrm{Cov}(X) < \mathrm{Cov}_{\max}(Z, C)$, because applying the same construction to such an optimizer would strictly improve the objective. Hence every optimizer attains maximal coverage, proving part (2). □

B.1.2 Proof of Theorem 3.2.

Theorem 3.2 (CNA maximizes a lower bound on matches). *Suppose Assumption 1. If allocation X^* is safe and maximizes **CNA**(C, u) for any increasing and strictly discretely concave u, then*

$$\mathbb{E}\left[\mathrm{Matches}_Z(X^*)\right] \geq L\left[\max_{X \text{ feasible for } Z,C} \mathbb{E}\left[\mathrm{Matches}_Z(X)\right]\right],$$

*i.e., **CNA** provides an L-approximation to our match objective function among feasible allocations.*

Proof. Recall that $k_i(X) = \sum_j X_{ij}$, the number of nudges that applicant i receives. Let $K_i = \sum_j Z_{ij}$ be the maximum number of nudges applicant i *can* receive, and recall that $q(k)$ is the probability that an applicant receiving k nudges adopts one.

Let $K = \max_i K_i$, let $\Delta_r = q(r) - q(r-1)$, and set $\Delta_{K+1} = 0$. For any feasible allocation X, summation by parts gives expected adoption as

$$\sum_i q(k_i(X)) = \sum_{r=1}^{K} \Delta_r \mathrm{Cov}_r(X) = \sum_{t=1}^{K} (\Delta_t - \Delta_{t+1}) \sum_{r=1}^{t} \mathrm{Cov}_r(X).$$

Discrete concavity makes every coefficient $\Delta_t - \Delta_{t+1}$ nonnegative. The coverage inequality established in the proof of Theorem B.1 therefore implies that X^* maximizes expected adoption among feasible allocations.

Under Assumption 1, every desirable match under any feasible X is made by an applicant who adopted a nudge, so

$$\mathbb{E}\left[\mathrm{Matches}_Z(X)\right] \leq \sum_i q(k_i(X)).$$

Applying the Oracle guarantee on acceptance rates in expectation for each program and summing over programs therefore lower-bounds desirable matches by L times expected adoption. Hence, for every feasible X,

$$\mathbb{E}\left[\mathrm{Matches}_Z(X^*)\right] \geq L \sum_i q(k_i(X^*)) \geq L \sum_i q(k_i(X)) \geq L\mathbb{E}\left[\mathrm{Matches}_Z(X)\right].$$

Taking the maximum over feasible X completes the proof. □

B.2 Proof of algorithm guarantees (Theorem 3.3)

Theorem 3.3 (Greatest safe nudge capacities). *Suppose Assumption 2. Then NudgeCapComputation$(O, \overline{c})$ terminates using $O\big(|\mathcal{P}|\left(1+\sum_{i,j} Z_{ij}\right)\log\left(1+\sum_{i,j} Z_{ij}\right)\big)$ oracle calls, and the returned nudge capacities C^* are safe. Moreover, for any safe $C \leq \overline{c}$, it holds that $C \leq C^*$ coordinatewise; thus, C^* is the unique coordinatewise greatest safe vector bounded by $\overline{c}$.*

Proof. We prove each of the three guarantees in turn. Note that the binary search procedure invoked by NudgeCapComputation to find the largest feasible value no greater than the current value of a program's nudge capacity always terminates since the search space is integral and finite (between 0 and C_j), the feasibility check is well-defined, and by Assumption 2 O_j is non-increasing in the candidate value. The feasible set is nonempty because the no-nudge allocation is safe, and increasing the other programs' nudge capacities cannot increase congestion at program j.

Finite termination. At every iteration, the algorithm either terminates or strictly decreases at least one coordinate of C. Thus $\sum_{j\in\mathcal{P}} C_j$ strictly decreases by at least 1 at each nonterminal iteration. Since $C_j \geq 0$ for all $j \in \mathcal{P}$, the total sum is bounded below by 0. Therefore the algorithm terminates after at most $1 + \sum_{j\in\mathcal{P}} \overline{c}_j = 1 + \sum_{i,j} Z_{ij}$ iterations, including the final unchanged iteration. Each iteration uses $O(\log(1+\overline{c}_j))$ oracle calls for binary search at program j, giving $O\big(\big(1+\sum_{j\in\mathcal{P}} \overline{c}_j\big) \sum_{j\in\mathcal{P}} \log(1+\overline{c}_j)\big)$ oracle calls in total. Since $\overline{c}_j \leq \sum_{i,j} Z_{ij}$ for every j, this gives the stated bound.

Safe nudge capacities. At termination, $C^{(t)} = C^{(t-1)}$. For each program j, the update rule therefore verifies $O_j(\mathbf{CNA}(C)) = 1$ at the final nudge capacity vector. Thus, the constraints are satisfied at termination.

Greatest safe nudge capacities. Let C be the final nudge capacities at termination. Let $C' \leq \overline{c}$ be any nudge capacity vector satisfying the congestion constraints. We show inductively that $C' \leq C^{(t)}$. This holds initially because $C^{(0)} = \overline{c}$. Suppose it holds at iteration $t-1$. For each program j, C'_j is feasible in program j's update: it is no greater than $C_j^{(t-1)}$, and cross-program monotonicity implies that raising the other nudge capacities from C'_{-j} to $C_{-j}^{(t-1)}$ cannot cause program j's oracle outcome to change from satisfied to violated (intuitively, increasing capacity elsewhere cannot increase congestion at program j, by assumption). The update therefore chooses $C_j^{(t)} \geq C'_j$. Hence $C' \leq C^{(t)}$, and thus $C' \leq C$ at termination. □

AI Disclosure

An LLM (GPT 5.2–5.6) was used as a proof assistant, such as to help draw explanatory figures, explain ideas, and formalize proofs. The authors came up with all primary ideas and stand by all statements, and nothing was used without validation and editing. AI was not used for data analysis, and did not have any access to any non-public data.

C SUPPLEMENTARY SIMULATION DETAILS

C.1 Non-personalized recommendations

In all of our main text discussion, the **Naive** approach refers to nudging students to every high-performing program where they are nudge-eligible (they have *High* offer likelihood and the program is within 45 minutes of their home). Here, we compare this to the *non-personalized* approach used in prior interventions, where *every* student at each middle school receives the same list of all nearby, high-performing programs (i.e., recommendations do not incorporate individual offer likelihood), showing that even **Naive** personalization improves substantively over the non-personalized approach.

As we discuss in the introduction, non-personalized interventions must either (a) be restricted to programs that most students would be admitted to (e.g., historically undersubscribed programs), or (b) risk steering many students to programs where they are ultimately rejected. Under our definition of high-performing programs, option (a) would have severely limited the scale of the intervention. Only 28 of 56 high-performing programs had any open capacity in the 2025 admissions cycle

(without any nudges). Further, the number of high-performing programs where at least half of all students at target middle schools have *High* offer likelihood estimates is only 4 of 56. To quantify the consequences of retaining these programs (option (b)), we simulate the allocation method of nudging all students at target middle schools to all high-performing programs within 45 minutes of their middle school and evaluate outcomes under 20% adoption (following the same procedure as in the main text); without personalization, 41 of 56 high-performing programs reject *over 50%* of nudged applicants, as opposed to 5 of 56 under our **Naive** allocation method (and, of course, 0 of 56 under our **Balanced Capacitated** allocation method).

C.2 Displacement effects

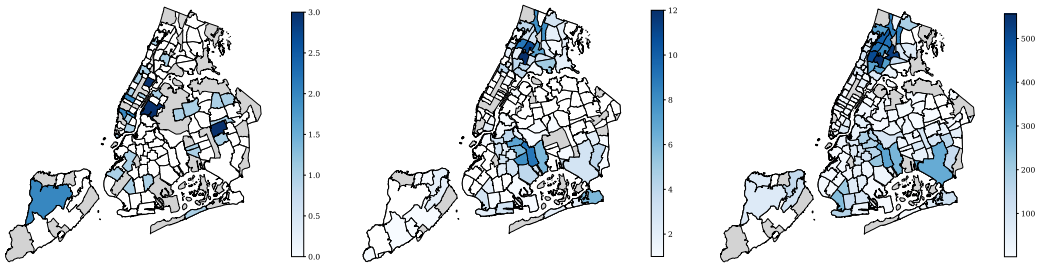
Our main congestion criterion in the deployment is preserving offer likelihoods for *High* offer likelihood students. In particular, if a student is told by NYCPS that their chance of admission at a program is over 99%, these predictions must remain reliable even under large-scale adoption of nudges. Here, we also analyze calibration of *Medium* and *Low* offer likelihoods, as well as the displacement effects across *all* students, based on simulations using historical 2024-25 application cycle data. We simulate 20% adoption under our **Balanced Capacitated** nudge allocation.

For *Medium* offer likelihood students, acceptance rates in simulation decline under nudge adoption but largely remain calibrated. For *Medium* offer likelihood students, the mean acceptance rate across nudge-eligible programs decreases from 63.8% to 49.0% (median: 69.2% to 52.1%), still well within the (1, 99)% stated interval. Of the 53 nudge-eligible programs that assign *Medium* offer likelihood to any students (of 56 total), 46 are calibrated in the absence of nudges. In our simulations at 20% adoption of nudges, acceptance rates for *Medium* offer likelihood students drop below the 1% threshold for 2 programs. Meanwhile, the mean nudge-eligible program acceptance rate of *Low* offer likelihood students changed from 8.36% to 5.60% (median 0.3% to 0.0%).

We next examine which students were displaced from nudge-eligible programs or newly match to nudge-eligible programs in the simulations, and we compare demographics between the groups. Among displaced students, 13.0% attend target middle schools. This indicates that the displacements increase representation of students from target middle schools at nudge-eligible high schools (and in particular of students with *High* offer likelihoods at these programs, indicating concordance with the program- or city-level desiderata). Increasing the number of these matches is in line with our definition of "desirable matches" used throughout the paper. Further, 75.7% of displaced students are FRL-eligible, compared to 86.1% of students who newly match to one these programs due to nudges.

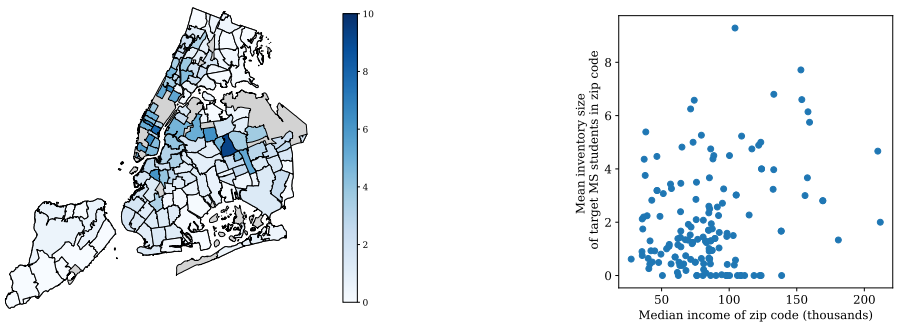
Overall, while nudges naturally induce some displacement, calibration is largely preserved and the resulting reallocation aligns with our objectives.

C.3 Other tables and figures



(a) **Nudge-eligible high schools** (b) **Target middle schools** (c) **Nudge-eligible students**

Fig. 8. **Number of intervention schools and students by zip code.** Gray zip codes are locations where no high school, middle school, or student (respectively) is located at all.



(a) Mean inventory size by zip code. Gray zip codes have ≤ 1 student who attend a target middle school.

(b) Zip code median income [Schroeder et al., 2025] vs. mean inventory size. Pearson's r = 0.28.

Fig. 9. **Disparities in access to nudge-eligible high schools.** A student's *inventory size* is number of high-performing programs they are nudge-eligible for. We compute the mean inventory size of students living in each zip code (among students attending target middle schools).

Allocation Method	Total # Nudges	# Students Nudged	Match Rate	# Matched to Nudged	# Matched to any Nudge-eligible
Naive	26,325	11,417	0.45	2,589	2,862
Capacitated	7,022	4,608	0.92	2,130	2,521
Balanced Capacitated	7,021	6,227	0.80	2,515	2,853

Table 2. **Simulation results at 50% student adoption.** At 50% adoption, enforcing nudge capacities continues to improve match rates with minimal effect on the total number of students matched to nudged and nudge-eligible programs.

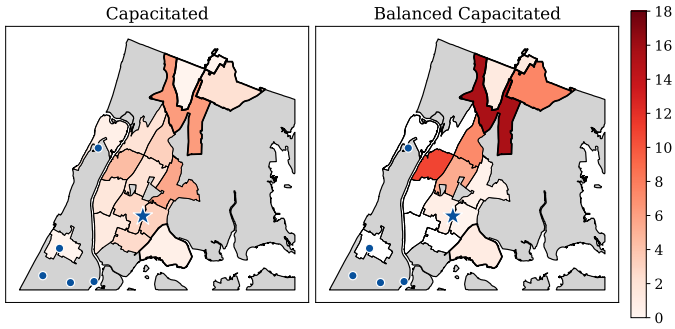
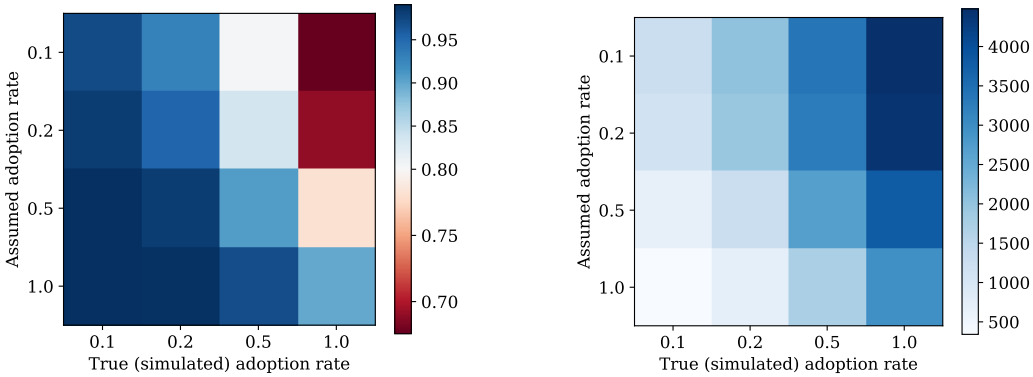


Fig. 10. **Geographic shift in nudged students.** The shade of red reflects the number of students from each zip code who are *matched* to Bronx Latin after adopting a nudge for the program, under each allocation method. Nudge-eligible programs are indicated by the blue star (Bronx Latin) and dots (other). Balanced allocation results in more matches from students in zip codes *without* access to any other programs (indicated with thicker outline).



(a) Program mean acceptance rate among *High* offer likelihood students.

(b) Total number of students matched to a nudged program.

Fig. 11. **Results of miscalibrated nudge capacities.** We sweep over pairs of assumed adoption rate (which informs nudge capacities) and true adoption rate, to characterize the behavior of miscalibrated nudge capacities. We find that setting nudge capacities too high or low navigates a tradeoff between *High* offer likelihood acceptance rates and total intervention effectiveness (number of matches to nudged programs).

D SUPPLEMENTARY DEPLOYMENT DETAILS

D.1 Details on NYC High School Match

New York City has about 400 high schools, attended by about 300,000 students. Each high school has at least one *program*; different programs at the same school may provide different concentrations (such as arts). There are about 900 total programs, and students apply to and are admitted at the *program level*. Before entering high school, students (often with the help of their families and school guidance counselors) create a ranked list of programs they wish to apply to, in an online system called MySchools. Since the 2024-25 admissions cycle, students have been able to list an

unlimited number of programs on their application profile. Programs vary in admissions standards, and each program also has its own (publicly available) criteria for ranking students, with different seat availability, eligibility, and priority groups. We note that applications for General Education (GE) and Students with Disabilities (SWD) are run through the same process but have separate seat groups at each program. Once a student opens their application in MySchools, they are assigned and able to view their lottery number, which is used to break ties between students in the same priority group. After the admissions process closes in December, students are then matched to a single program on their list using student-proposing deferred acceptance, run by the NYC Public Schools (NYCPS). If a student fails to match to any program on their list, NYCPS assigns them a manual placement based on other programs they applied to and distance from home.

The 2024-25 admissions cycle also saw the introduction of a new offer likelihood tool for students to use during the application process [Corradini and Idoux, 2025a]. The tool uses a bootstrapping and simulation approach to calculate whether a student has a *High* (>99%), *Medium*, or *Low* (<1%) chance of admission at a given program if they apply.[24] Since offer likelihood predictions are dependent on an individual student's lottery number, they are calculated and made visible to a student immediately after they begin an application in their MySchools portal. We directly use these predictions in our work.

Note that all students have access to these offer likelihood estimates, including control students and those not eligible for the intervention. Thus, our intervention does not provide any new information (offer likelihoods, program information) to students that is not already available to them through other sources (MySchools, commute time calculators)—it simply highlights certain programs and their characteristics.

D.2 Data

High school metric data. To determine which high school programs to involve in the intervention, we used public data characterizing schools through quantitative performance-based and survey-based metrics: graduation rates, college & career readiness scores, and school safety ratings. We used metrics reported on the 2023-24 school year, the most recent results available at the time of the deployment. All of these metrics are reported at the *school* level (even though applications are at the *program* level), so the performance of all programs at a school are assumed to be the same. In practice, according to NYCPS collaborators, students within a school often have access to the same resources across programs, so considering school-level metrics are reasonable proxies for program-level metrics. We also use student-level graduation data from the 2017-18 and 2018-19 application cycles to compute program-level graduation rates and find that 91% of programs at schools with school-level graduation rates above 95% (as reported in the 2023-24 metric data) have a computed program-level graduation rate of at least 90% across these two years.

Middle school feeder pattern data. To determine which middle schools to involve in the intervention, we used "feeder pattern" data from the 2023-24 and 2024-25 admissions cycles, characterizing how many students from each middle school enroll at each high school.

Student-level data. We have access to each student's entire application, including current school, ranked application profile of program preferences, match outcome, home census tract, lottery number (used for breaking ties within an admissions priority group), and offer likelihoods using the above-described tool. We consider both General Education (GE) and Students with Disabilities (SWD), which use the same match process but have separate seat groups at each program. We

[24]Not all programs have offer likelihoods available. For example, if a program uses qualitative/human graded information such as auditions/essays, then offer likelihoods are not available.

developed the system based on analysis and simulations using historical data from the 2024-25 admissions cycle (72,197 students), i.e., we calculate program-level nudge capacities using 2024-25 data. Simulations on 2024-25 admissions cycle data use a student's zip code instead of census tract due to data availability constraints. We deploy our system for the 2025-26 admissions cycle, in which we had access to lottery numbers on a rolling basis, as students began applications.

Match simulator. Lastly, we have access to details on each program's capacity and criteria for ranking students for the 2024-25 admissions cycle. We used this data to build a simulator that runs the deferred acceptance algorithm using these settings, meaning that we can accurately simulate match outcomes under any assumptions on student applications for this application cycle.

D.3 Balance checks

We assess balance in student characteristics across treatment groups. While the realized groups appear to differ somewhat in residence borough and race composition, the distribution of p-values from corresponding balance tests is approximately uniform, consistent with valid randomization. This suggests that observed differences reflect random variation in the realized assignment (rather than systematic imbalance). To account for observed imbalances, we control for these characteristics in regressions as supplementary analyses, as shown in Tables 6 and 7.

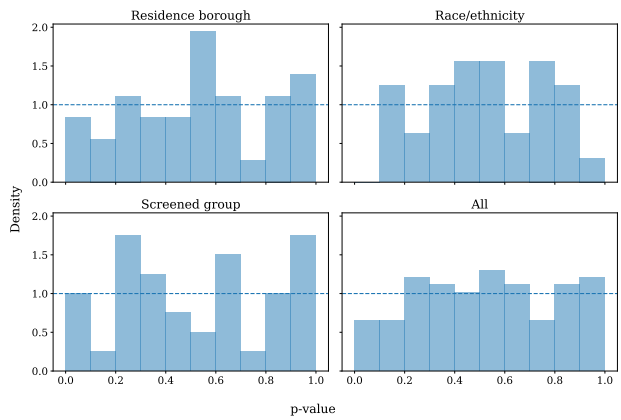


Fig. 12. **Distribution of p-values from balance tests.** For each covariate category (e.g., residence borough = Queens), we use Y as an indicator for the category and estimate linear probability models of the form $Y \sim$ treatment_intensity_50 + treatment_intensity_100. We estimate these models using (i) all students at target middle schools and (ii) only nudge-eligible students at those schools. The figure plots the distribution of p-values from these tests. The approximately uniform distribution is consistent with differences arising from random variation under valid random assignment.

Variable	Treatment	Control	Non-target
Total #	111	107	636
Total # students	9,925	11,593	39,062
Bronx	0.41	0.41	0.19
Brooklyn	0.37	0.38	0.27
Manhattan	0.10	0.09	0.27
Queens	0.10	0.09	0.25
Staten Island	0.03	0.02	0.03
Asian	0.060	0.049	0.277
Black	0.310	0.324	0.112
Hispanic/Latino	0.525	0.521	0.382
White	0.075	0.079	0.187
SWD	0.226	0.228	0.171
ELL	0.223	0.208	0.185
Mean # nudge-eligible students	6.9	7.5	—
Median # nudge-eligible students	4	5	—
Mean grade tier	3.33	3.35	2.91

Table 3. **Middle school characteristics.** Target middle schools (distributed within treatment and control) are those that historically did not send many students to high-performing high schools, so we expect them to systematically differ from middle schools not in the experimental group. Non-target middle schools reported here exclude charter schools (which were also filtered out for computing target middle schools); students from private middle schools and homeschool further contribute to the discrepancy between the number of students reported here and the total number of students in the match.

Variable	**Treatment**	**Control**
Total #	444	822
Bronx	0.23	0.39
Brooklyn	0.48	0.41
Manhattan	0.07	0.09
Queens	0.22	0.11
Staten Island	0.00	0.00
Asian	0.072	0.063
Black	0.275	0.315
Hispanic/Latino	0.517	0.542
White	0.086	0.055
Female	0.538	0.491
SWD	0.187	0.223
ELL	0.169	0.181
Screened Group 1	0.30	0.27
Screened Group 2	0.18	0.20
Screened Group 3	0.21	0.22
Screened Group 4	0.16	0.14
Screened Group 5	0.16	0.16
Mean screened group	2.7	2.7

Table 4. **Nudge-eligible student characteristics.** Note that there are some imbalances between student demographic distributions between treatment and control, especially with respect to Borough. However, these do not explain our results: our primary regression specifications contain stratum fixed effects, which are defined by middle school Borough and number of nudge-eligible students in the 2024-25 admissions cycle.

Variable	**Treatment**	**Control**
Student-level measures (per student)		
Mean # in inventory	1.65	1.42
Median # in inventory	1	1
Mean # nudged	1.60	1.35
Median # nudged	1	1
Distances to nudged programs (minutes, per student–nudge pair)		
Mean travel time to nudged program	37.07	38.12
Median travel time to nudged program	39.22	39.89
Program performance among nudged programs (per student–nudge pair)		
Mean graduation rate of program	96.97	97.35
Median graduation rate of program	96.4	98.0
Mean CCR of program	70.5	70.2
Median CCR of program	70.4	70.4
Geographic distribution of nudged programs (share of student–nudge pairs)		
Bronx	0	0
Brooklyn	0.15	0.14
Manhattan	0.27	0.42
Queens	0.59	0.45
Staten Island	0	0

Table 5. **Breakdown of characteristics on nudge-eligible students and nudged programs.**

D.4 Treatment effects with fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Treatment effect	0.553	0.548	0.543	0.546	0.560	–	0.552	0.557	0.497	0.493	0.469	0.472	0.523	0.518	0.517	0.521
	(0.239)	(0.240)	(0.235)	(0.236)	(0.230)	–	(0.225)	(0.227)	(0.249)	(0.250)	(0.245)	(0.246)	(0.237)	(0.238)	(0.230)	(0.232)
p-value	0.021**	0.022**	0.021**	0.021**	0.015**	–	0.014**	0.014**	0.046**	0.049**	0.055*	0.055*	0.027**	0.029**	0.025**	0.024**
Residence borough FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes								
Ethnicity FE	Yes	Yes	Yes	Yes					Yes	Yes	Yes	Yes				
Screened group FE	Yes	Yes			Yes	Yes			Yes	Yes			Yes	Yes		
Gender FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	

Table 6. **Treatment effects with alternative fixed effects specifications (Logit).** *Note:* Maximum-likelihood estimation for specification (6) failed to converge. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Treatment effect	0.059	0.058	0.061	0.061	0.060	0.060	0.063	0.064	0.052	0.052	0.053	0.053	0.056	0.056	0.059	0.060
	(0.026)	(0.026)	(0.027)	(0.027)	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)	(0.028)	(0.029)	(0.029)	(0.027)	(0.027)	(0.028)	(0.028)
p-value	0.024**	0.026**	0.025**	0.025**	0.019**	0.021**	0.018**	0.018**	0.058*	0.061*	0.068*	0.068*	0.036**	0.038**	0.034**	0.034**
Residence borough FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes								
Ethnicity FE	Yes	Yes	Yes	Yes					Yes	Yes	Yes	Yes				
Screened group FE	Yes	Yes			Yes	Yes			Yes	Yes			Yes	Yes		
Gender FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
R^2	0.083	0.083	0.033	0.033	0.072	0.072	0.009	0.009	0.081	0.081	0.029	0.029	0.071	0.071	0.008	0.007

Table 7. **Treatment effects with alternative fixed effects specifications (OLS).** Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Analysis	Status
Primary experimental outcomes	
Primary outcome: applications	Logit results in Figure 5b; OLS results in Table 10
Primary outcome: matches	Logit results in Figure 5b; OLS results in Table 10
Alternative primary outcome specifications	Table 10
Secondary experimental outcomes	
Scale-up analyses	Section 6.3
Click outcomes	Section 6.2
Survey outcomes	
Survey primary outcome	Table 16
Survey secondary outcomes	Table 17
Survey exploratory outcomes	Tables 18 and 19
Survey qualitative outcomes	Appendix D.7
Further exploratory analyses	
Local average treatment effects	Table 12
Heterogeneous treatment effects	Table 13
Quantitative metrics of matched/enrolled school	Matched school only (enrollment data not yet available): Table 15
Within-applicant comparisons of application profiles over time	Appendix D.6
Outcomes at school level	Not reported; unable to compute full set of nudged students in deployment year due to data limitations
Long-term outcomes	Not yet realized

Table 8. **Pre-registration compliance for analyses.** This table summarizes the pre-registered analyses reported, along with deviations from the pre-registered plan.

D.5 Pre-registration compliance

In this section, we present all procedures and analyses detailed in the pre-registration, and we describe the status of each. Nudge emails were implemented as described, with the exception that students from continuing middle schools were dropped from Batch 2, in response to stakeholder feedback. The same rule (on a feature independent from original randomization) was applied to control and treatment students. The pre-registration also stated that we would send follow-up emails to students a few weeks after their first intervention email. Due to the timing of original intervention emails (which faced a variety of unforeseen logistical constraints), follow-up emails were sent for Batch 1 students only. We also excluded students from receiving continuing middle schools from follow-up emails. Lastly, we pre-registered that we would send text messages as an additional method of communication to treatment families; this was not implemented, due to unforeseen issues with NYCPS text message platform. All other aspects of the intervention itself are in full compliance with the pre-registration. For analyses, we detail pre-registration compliance in Table 8.

D.6 Additional Experimental Results

In this section, we report the remaining pre-registered experimental outcomes. We also provide raw rates of applications and matches to nudged and nudge-eligible programs in Figure 13. "Nudge-eligible programs" refers to the 15 deployed nudge-eligible programs included in our experiment this year; this is distinct from nudge-eligible student-program pairs (desirable matches), which indicate that a student meets the *High* offer likelihood, commute distance, and middle school feeder pattern requirements at that specific program.

Regression robustness checks. Our pre-registration specified that we would measure treatment effects for the primary outcome through logistic regression. For robustness, we further report results using linear probability models (Table 10) to show that all significance results remain unchanged, with the exception of "Ranked any nudged program in top 3" without stratum fixed effects, which changes from being significant at the $p < 0.10$ level to $p=0.111$ with OLS.

Additional outcome specifications. We report results for additional outcome specifications in Table 9 (logistic regression) and Table 10 (linear regression). First, we measure treatment effects on whether applicants ranked any *nudge-eligible* program, rather than only the programs they were individually nudged to. While these treatment effects are not consistently significant across specifications (with vs. without stratum fixed effects), the behavioral changes in applying to nudged programs appears to be additive, translating to similar percentage-point increases in nudge-eligible programs. We also report results for the two pre-registered alternative specifications of our primary outcome: whether an applicant ranked a nudged program first, and the rank position of the first nudged program on the applicant's list (conditional on ranking any nudged program).[25] We do not find a statistically significant treatment effect under either specification.

Exploratory spillover analysis. We fit a linear regression with middle school treatment-intensity coefficients (Table 11). Untreated students in 50% treatment-intensity target middle schools exhibited higher application rates to nudged programs than students in 0% schools. While this pattern is consistent with possible within-school spillovers, it may also reflect residual school-level imbalance or noise, since the spillover estimate is identified from between-school variation and a relatively small number of untreated students in partially treated schools.

Local Average Treatment Effects. As preregistered, we estimate local average treatment effects (LATE) using treatment assignment as an instrument for email engagement, defined as clicking at least one tracked link. By this definition, 19.8% of treatment students complied with the intervention. However, our results indicate that many students likely read and engaged with the email without clicking on any link, and so the exclusion restriction is violated—these effects are likely overstated. While we report this analysis in Table 12 to be consistent with the preregistration, we do not rely on them.

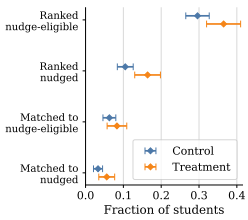
Heterogeneous treatment effects. For each subgroup variable (FRL/ELL/STH status, residence borough, screened group, race/ethnicity, gender), we estimate a regression with an interaction coefficient between treatment assignment and the feature: $Y_i \sim$ treatment*feature. We report results in Table 13. We find no evidence of heterogeneous treatment effects, though our RCT was not powered to detect such effects.

Additional analysis of link-clicking behavior. We find that the mean number of link clicks per treated applicant is 0.502, and on average applicants click on links for 18.2% of their nudged programs. We also examine whether link click behaviors vary across subgroups, regressing an

[25]Because rank positions are non-binary, we estimate that specification using linear models only.

indicator for clicking any link on demographic characteristics among treated applicants. Overall, we find limited evidence of heterogeneous engagement, although we were underpowered to do so (Table 14).

Comparisons between pre-intervention and submitted application lists. We also examine saved student application lists from before the intervention, to study how and whether applicants revise their existing lists in response to nudges. We find that only 10 of 444 treatment applicants had any saved programs as of the data pulls one day before their nudge email was sent (this number is 18 of 822 for control applicants). None of these 10 applicants had a nudged program on their pre-intervention saved list, and none added a nudged program to their final application. This low rate of adding schools to the portal is consistent with our intervention only targeting applicants with in-progress, un-submitted applications—but given the small and potentially biased sample of applicants who had already added schools to their application upon receiving the email, means we have limited ability to study student behavior in adding schools to an existing list.



				Ranked program…				
Group	**Total**	**Submitted**	**Program type**	*Anywhere*	*First*	*Second*	*Third*	**Matched**
Control	822	803 (97.7%)	*Nudged*	86 (10.5%)	19 (2.3%)	11 (1.3%)	11 (1.3%)	27 (3.3%)
			Nudge-eligible	243 (29.6%)	53 (6.4%)	36 (4.4%)	29 (3.5%)	52 (6.3%)
Treatment	444	439 (98.9%)	*Nudged*	73 (16.4%)	17 (3.8%)	9 (2.0%)	9 (2.0%)	25 (5.6%)
			Nudge-eligible	162 (36.5%)	36 (8.1%)	21 (4.7%)	19 (4.3%)	37 (8.3%)

Fig. 13. **Summary of student behavior in response to nudges.** For both control and treatment groups, we summarize the number of students total, who ultimately submitted an application, and who ranked (anywhere on the list or in any of the first 3 spots) or matched to a nudged or nudge-eligible program in their application.

Table 9. **Regression-estimated treatment effects (Logit): additional outcome specifications**

Outcome	**Control mean**	**Treatment mean**	**Stratum FE**	**Treatment effect**	**p-value**
Ranked any nudge-eligible program	0.296	0.365	Yes	0.226 (0.139)	0.104
			No	0.314 (0.155)	0.042**
Ranked any nudged program first	0.023	0.038	Yes	0.638 (0.456)	0.162
			No	0.520 (0.418)	0.213
Ranked any nudged program in top 3	0.050	0.079	Yes	0.545 (0.298)	0.068*
			No	0.489 (0.292)	0.094*
Matched to any nudge-eligible program	0.063	0.083	Yes	0.279 (0.294)	0.344
			No	0.297 (0.276)	0.281

Logistic regressions. Robust standard errors clustered at the school level in parentheses. Treatment is randomly assigned within strata; regressions are estimated with and without stratum fixed effects (borough and school size). Control and treatment means are the average outcomes among control and treatment students in the experimental population, respectively. Coefficients are log-odds ratios for each outcome. There are 1,266 total observations. * $p < 0.10$, ** $p < 0.05$

Table 10. **Regression-estimated treatment effects (OLS): primary outcomes and additional outcome specifications**

Outcome	Control mean	Treatment mean	Stratum FE	Treatment effect	p-value
Ranked any nudge-eligible program	0.296	0.365	Yes	0.049 (0.030)	0.105
			No	0.069 (0.034)	0.041**
Ranked any nudged program	0.105	0.164	Yes	0.064 (0.026)	0.013**
			No	0.060 (0.028)	0.034**
Ranked any nudged program first	0.023	0.038	Yes	0.0185 (0.014)	0.188
			No	0.0152 (0.013)	0.244
Ranked any nudged program in top 3	0.050	0.079	Yes	0.032 (0.018)	0.079*
			No	0.029 (0.018)	0.111
Matched to any nudge-eligible program	0.063	0.083	Yes	0.019 (0.020)	0.353
			No	0.020 (0.019)	0.292
Matched to any nudged program	0.033	0.056	Yes	0.027 (0.016)	0.095*
			No	0.023 (0.016)	0.136
Rank position of first nudged program	3.91	4.16	Yes	-0.008 (0.604)	0.990
			No	0.257 (0.625)	0.681

Linear probability models. Robust standard errors clustered at the school level in parentheses. Treatment is randomly assigned within strata; regressions are estimated with and without stratum fixed effects (Borough and school size). Control and treatment means are the average outcomes among control and treatment students in the experimental population, respectively. When coefficients represent changes in the probability, multiplying by 100 gives the percentage-point effect. There are 1,266 total observations for all regressions except "Rank position of first nudged program", for which there are 159 total observations. * $p < 0.10$, ** $p < 0.05$

Table 11. **Spillover effects**

	Stratum FE	Coefficient	p-value
Treatment intensity 50	Yes	0.0837	0.044**
	No	0.0778	0.079*
Treatment intensity 100	Yes	0.0502	0.276
	No	0.0436	0.374
Treatment	Yes	0.0213	0.530
	No	0.0198	0.556

Linear probability model with robust standard errors clustered at the school level. * $p < 0.10$, ** $p < 0.05$

Table 12. **Local Average Treatment Effects for primary outcomes and alternative outcome specifications.** We define an observed complier as a treatment applicant who clicks on at least one tracked link in their intervention email. We note that these effects are likely overstated—more people read the email than clicked on a tracked link, so not all compliers were observed. We do not track email open rates directly.

Outcome	Control mean	Observed complier mean	Stratum FE	LATE	p-value
Ranked any nudge-eligible program	0.296	0.500	Yes	0.253 (0.144)	0.078*
			No	0.349 (0.138)	0.012**
Ranked any nudged program	0.105	0.284	Yes	0.333 (0.104)	0.001***
			No	0.302 (0.098)	0.002***
Ranked any nudged program first	0.023	0.091	Yes	0.096 (0.052)	0.065*
			No	0.077 (0.049)	0.120
Ranked any nudged program in top 3	0.050	0.159	Yes	0.167 (0.074)	0.024**
			No	0.146 (0.070)	0.038**
Matched to any nudge-eligible	0.063	0.136	Yes	0.098 (0.080)	0.220
			No	0.101 (0.076)	0.182
Matched to any nudged program	0.033	0.114	Yes	0.138 (0.062)	0.026**
			No	0.118 (0.059)	0.044**

Local average treatment effects estimated using two-stage least squares (2SLS) linear probability models, with randomized treatment assignment as an instrument for clicking at least one tracked link in the nudge email. Robust standard errors clustered at the school level in parentheses. Treatment is randomly assigned within strata; regressions are estimated with and without stratum fixed effects (Borough and school size). Control and observed complier means are the average outcomes among control students and observed compliers in the experimental population, respectively. Coefficients represent probability changes; multiplying by 100 gives the percentage-point effect. There are 1,266 total observations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Feature	**Coef.**	**p-value**
FRL/ELL/STH status		
Treatment (Non-FRL/ELL/STH)	0.0176	0.745
Treatment × FRL/ELL/STH	0.0499	0.336
Residence borough		
Treatment (Brooklyn)	0.053	0.186
Treatment × Manhattan	-0.021	0.794
Treatment × Queens	-0.055	0.430
Treatment × Bronx	0.076	0.259
Screened group		
Treatment (Group 1)	0.091	0.121
Treatment × Group 2	0.002	0.979
Treatment × Group 3	-0.092	0.119
Treatment × Group 4	-0.047	0.444
Treatment × Group 5	-0.046	0.492
Race/ethnicity		
Treatment (Hispanic)	0.039	0.266
Treatment × Asian	0.038	0.767
Treatment × Black	0.040	0.395
Treatment × White	-0.038	0.689
Treatment × Multi-racial	0.161	0.138
Gender		
Treatment (Female)	0.076	0.033**
Treatment × Male	-0.034	0.345

Table 13. **Heterogeneous treatment effects.** Outcome is whether the applicant applied to a nudged program.

Table 14. **Heterogeneity in link-clicking behavior among treated applicants**

	Any link click		**# link clicks**		**Frac. links clicked**	
Subgroup variable	*Coef.*	*p-value*	*Coef.*	*p-value*	*Coef.*	*p-value*
FRL/ELL/STH status						
FRL/ELL/STH	-0.0840	0.099	-0.143	0.424	-0.0739	0.147
Screened group						
Group 2	0.0525	0.335	-0.0423	0.802	0.0463	0.397
Group 3	0.0085	0.890	-0.0094	0.966	-0.0039	0.948
Group 4	0.0605	0.357	0.0782	0.750	0.0411	0.522
Group 5	0.0246	0.715	0.0468	0.840	0.0278	0.664
Race/ethnicity						
Black	-0.0559	0.428	-0.1034	0.497	-0.0684	0.333
Hispanic	-0.0085	0.905	0.1628	0.307	-0.0282	0.684
Multi-racial	0.2227	0.033	0.2549	0.171	0.2227	0.033
White	0.0310	0.722	0.2988	0.230	-0.0019	0.983

Each row reports the coefficient from a separate linear regression among treated applicants. FRL/ELL/STH indicates whether a student is classified as free/reduced-price lunch eligible, an English language learner, or a Student in Temporary Housing. The screened group (based on student grades) regressions use Group 1 as the reference category. The race/ethnicity regressions use Asian students as the reference category.

Table 15. **Treatment effect on matched school metrics.** Reported metrics are those used to define high-performing schools.

Metric	**Control mean**	**Stratum FE**	**Coef.**	**p-value**
Graduation Rate	90.67	Yes	0.27 (0.7)	0.675
		No	0.17 (0.8)	0.820
College and Career Readiness	49.78	Yes	1.79 (1.16)	0.122
		No	2.19 (1.38)	0.111
Safety	3.62	Yes	0.025 (0.047)	0.589
		No	-0.031 (0.070)	0.662

Linear models. Robust standard errors clustered at the school level in parentheses. Treatment is randomly assigned within strata; regressions are estimated with and without stratum fixed effects (Borough and school size). Control mean is the average outcome among control students in the experimental population. Metrics are not available for all schools, and matches to those schools are omitted; this results in 1,228 total observations for Graduation Rate and College and Career Readiness regressions, and 1,191 for Safety regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D.7 Additional survey results

We analyze post-application survey responses to measure whether the nudge emails increased applicants' familiarity with high-performing high school programs near their home, as pre-registered. We sent the survey to all treatment and control students from middle schools retained over both batches (370 and 707, respectively), and an equal number of non-treatment students, for a total of 2,154 total students.

Our survey had 42 treatment respondents, 61 control respondents, and 136 respondents from non-target middle schools. For each student, we sample one nearby, high-performing high school program to ask more detailed questions about; for treatment and control students this program is randomly sampled from nudged programs they *applied to* if any exist, and from all nudged programs otherwise. For non-target students, the program is randomly sampled from all nearby, high-performing high school programs.

Table 16 summarizes the primary survey outcomes (distinct from primary experimental outcomes): on the survey sample, we do not find evidence of treatment effect on self-reported familiarity or rates of recognition on sampled programs. Table 17 details all other survey responses, reporting raw survey rates for other questions. Due to small sample size, these results should be interpreted as descriptive evidence, potentially providing insight for future interventions.

We also report outcomes specifically among students who report reading the email (Table 18) and run exploratory regressions by applicants' self-reported primary information sources for high school options (Table 19).

Finally, we conducted a thematic analysis of open-ended survey responses. Among families who answered why they did not apply to the sampled high-performing program, most responses referred either to the program's specific academic interest area or the fact that it was too far from home. Only 2 of 25 families provided a free-response answer regarding their perception of the email, so we do not draw conclusive insights from this question. In response to "what advice would have helped you navigate the application process better", the most common themes were "nothing", more school/program information, insight on admissions chances, personalized recommendations, clarity on the process and admissions requirements, and better support for different languages or students with individualized education programs.

Outcome	**Model**	**Coefficient**	p**-value**
Self-reported familiarity with high-performing programs near home (very unfamiliar (1) – very familiar (4) scale)	Ordered logit	-0.25	0.517
	OLS	-0.11	0.590
Familiar with sampled high-performing program	Logit	0.368	0.416
	OLS	0.089	0.409

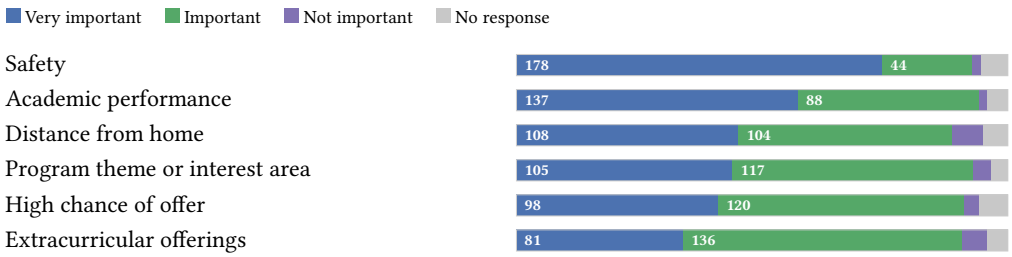
Table 16. **Primary survey outcomes.** Familiarity with high-performing programs (self-reported familiarity and rates of recognition on sampled programs), comparing treatment and control respondents. Ordered logit and logit coefficients are reported in log-odds units. For self-reported familiarity, the OLS coefficient is measured in points on the 1–4 scale. For the binary familiarity outcome, the OLS coefficient can be interpreted as a change in probability; multiplying by 100 gives the percentage-point effect. 103 total observations.

D.8 Scale-up analyses

Table 17. **Descriptive survey outcomes.** We report raw survey rates for other outcomes. These are exploratory and could inform future interventions.

Outcome	Treatment	Control	Non-target
Panel A. Reaction to email (Treatment group only)			
Read the email	59.5% (25/42)	–	–
Email was helpful (of applicants who read email)	88.0% (22/25)	–	–
Email affected application decisions (of applicants who read email)	44.0% (11/25)	–	–
Perception of sampled program improved after reading (of applicants who read email and recognized sampled prog)	66.7% (10/15)	–	–
Would *not* have applied *without* the email (of applicants who applied to sampled prog)	72.7% (8/11)	–	–
Panel B. Knowledge about programs (Treatment & control groups)			
Recognized sampled program	52.4% (22/42)	45.9% (28/61)	–
Email readers: knowledge *before* email			
Sampled prog is at high-performing school	33.3% (5/15)	–	–
High likelihood at sampled prog	13.3% (2/15)	–	–
Sampled prog is within 45 min of home	20.0% (3/15)	–	–
Non-email readers: knowledge at end of process			
Sampled prog is at high-performing school	85.7% (6/7)	89.3% (25/28)	–
High likelihood at sampled prog	71.4% (5/7)	57.1% (16/28)	–
Panel C. Demand for recommendations (Control & non-target groups)			
Would have considered applying to sampled prog	–	64.8% (35/54)	56.9% (62/109)
Interested in receiving recommendations	–	90.2% (55/61)	88.1% (118/134)
Panel D. Reasons for not applying to sampled program (All groups)			
Multiple choice among applicants who recognized sampled prog but did not apply			
Did not know enough about it	22.2% (2/9)	31.8% (7/22)	21.4 % (9/42)
Did not match child's interests	44.4% (4/9)	31.8% (7/22)	57.1% (24/42)
Did not think child would get in	22.2% (2/9)	9.1% (2/22)	16.7% (7/42)
Too far from home	22.2% (2/9)	9.1% (2/22)	4.8% (2/42)

Panel E. Feature importance for recommendations (Aggregated across all groups)



The treatment, control, and non-target columns report raw survey outcomes. The non-target group is included for descriptive context only and is not part of the pre-registered treatment-control comparison. Some exploratory outcomes are conditional on responding to follow-up questions, so denominators vary across rows.

Table 18. **Outcomes by applicants' self-reported email readership (exploratory)**

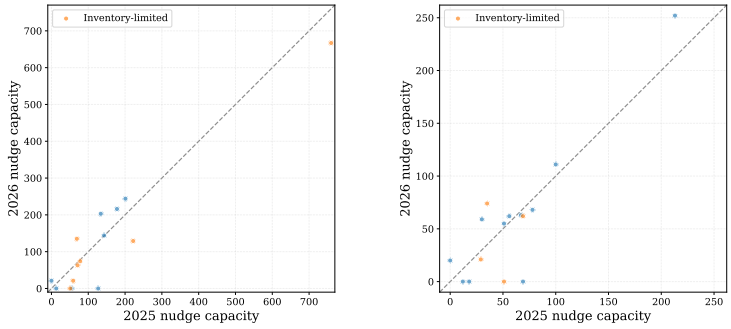
Outcome	Read email	Did not read email
Self-reported familiarity with high-performing programs near home (very unfamiliar (1) – very familiar (4) scale)	2.72 ± 0.37	3.06 ± 0.52
Familiar with sampled high-performing program	60.0% (15/25)	41.2% (7/17)
Applied to at least one nudged program	44.0% (11/25)	29.4% (5/17)
Applied to at least one nudge-eligible program	64.0% (16/25)	58.8% (10/17)

This table conditions on treatment assignment and splits respondents by whether they self-reported reading the email. Results for self-reported familiarity include 95% confidence intervals.

Table 19. **Regressions for outcomes by applicants' primary information sources (exploratory)**

Info source	Familiarity	Read email	Applied to nudged prog	Interested in recs
Intercept	2.930 (0.119)	0.449 (0.135)	0.156 (0.045)	0.853 (0.056)
Other parents	-0.047 (0.124)	-0.246 (0.170)	-0.071 (0.035)	-0.033 (0.052)
Paid consultants	-0.018 (0.472)	-0.764 (0.204)	-0.051 (0.033)	0.126 (0.076)
Counselors	-0.016 (0.104)	0.162 (0.137)	-0.006 (0.037)	0.026 (0.050)
MySchools	0.032 (0.104)	-0.066 (0.169)	-0.035 (0.037)	-0.005 (0.054)
Other websites	0.062 (0.167)	-0.180 (0.218)	-0.038 (0.040)	-0.054 (0.076)
Information sessions	0.077 (0.122)	0.399 (0.114)	-0.017 (0.041)	0.082 (0.043)
Observations	239	42	239	197

This table reports coefficients from exploratory regressions with indicators for applicants' self-reported primary information sources. Outcomes are: self-reported familiarity with high-performing programs near home on a 1–4 scale (all respondents); whether the applicant reported reading the email (treatment respondents only); whether the applicant applied to at least one nudged program (all respondents); and whether the applicant reported interest in receiving recommendations (control and non-target respondents only). Standard errors are reported in parentheses. These regressions are descriptive because information sources are self-reported and not randomly assigned.



(a) **40% nudge adoption** (b) **100% nudge adoption**

Fig. 14. **Comparison of nudge capacities between years.** We compute nudge capacities on 2024-25 versus 2025-26 application data, assuming two simulated adoption rates. "Inventory-limited" orange dots indicate programs were limited by the number of nudge-eligible students rather than congestion.

D.9 Other figures

Subject: Important Information About Your Child's High School Options

Dear [parent name],

Your child, [student name] ([student ID]), has been selected to receive a personalized list of programs at high-performing high schools, where they have a high chance of receiving an offer! All are within 45 minutes of your home. These schools were selected based on their graduation rates, college & career readiness scores, and safety metrics.

Your chance of receiving an offer at each program is based on data from previous years, including offer rates and application data. Although we **cannot guarantee admission**, we encourage you to explore these programs and learn about the opportunities they offer!

[school 1 name] ([school 1 code])
[school 1 address]

- [school 1 program 1 name] ([school 1 program 1 code])
- [school 1 program 2 name] ([school 1 program 2 code])
- …

[school 2 name] ([school 2 code])
[school 2 address]

- [school 2 program 1 name] ([school 2 program 1 code])
- [school 2 program 2 name] ([school 2 program 2 code])
- …

…

If you have any questions, please email HSenrollment@schools.nyc.gov.

Best,

The High School Admissions Team
NYC Public Schools

Fig. 15. **Nudge email text.**