ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-396

Choosing Alone? Peer Similarity in High School Choices

Nicholas D.E. Mark
New York University

Sean Corcoran
Vanderbilt University

Jennifer Jennings
Princeton University

We provide a descriptive analysis of within-school and neighborhood similarity in high school applications in New York City. We depart from prior work by examining similarity in applications to specific schools rather than preferences for school characteristics. We find surprisingly low similarity within schools and neighborhoods, but substantial variation by race and prior achievement. White and Asian students are more likely to have choices in common relative to Black and Hispanic students, a difference that persists after controlling for achievement and location. Likewise, higher-achieving students are more likely to have choices in common, conditional on other student characteristics and location. An implication is that students' likelihood of attending high school without any peers from their middle school or neighborhood varies by student background.

VERSION: May 2021

Suggested citation: Mark, Nicholas D.E., Sean P. Corcoran, and Jennifer L. Jennings. (2021). Choosing Alone? Peer Similarity in High School Choices. (EdWorkingPaper: 21-396). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/d1z0-tj58

Choosing Alone? Peer Similarity in High School Choices

Nicholas D.E. Mark[1]
New York University

Sean Corcoran
Vanderbilt University

Jennifer Jennings
Princeton University

May 4, 2021

Abstract: We provide a descriptive analysis of within-school and neighborhood similarity in high school applications in New York City. We depart from prior work by examining similarity in applications to *specific* schools rather than preferences for school characteristics. We find surprisingly low similarity within schools and neighborhoods, but substantial variation by race and prior achievement. White and Asian students are more likely to have choices in common relative to Black and Hispanic students, a difference that persists after controlling for achievement and location. Likewise, higher-achieving students are more likely to have choices in common, conditional on other student characteristics and location. An implication is that students' likelihood of attending high school without any peers from their middle school or neighborhood varies by student background.

Acknowledgements: We would like to thank the New York City Department of Education, James Kemple, and the Research Alliance for New York City Schools for providing access to the data required for this study. We benefited greatly from feedback from Carolyn Sattin-Bajaj, Lauren Covelli, conference participants at the 2019 Association for Education Finance and Policy Annual Conference, and research staff members at the New York City Department of Education.

[1] Corresponding author. Department of Sociology, New York University, Puck Building 4th floor, 295 Lafayette Street, New York, NY 10012-9605. Email: nm2648@nyu.edu

An important body of school choice research has estimated the implicit "weights" families place on characteristics such as academic performance, peer composition, and distance when choosing schools (Abdulkadiroğlu et al., 2020; Glazerman & Dotter, 2017; Harris & Larsen, 2015; Hastings, Kane, & Staiger, 2006). That these weights vary by race/ethnicity and socioeconomic status raises the question of what social processes contribute to this variation. On the one hand, residential segregation combined with a preference for proximity imply a common *supply* of school options for families with shared characteristics (Edwards, forthcoming). In systems with academic screening, a student's prior achievement also expands or limits their choice set, and achievement is correlated with family resources (e.g., Sartain & Barrow, 2020). At the same time, *preferences* for school characteristics can vary by family background, and students in the same neighborhood, school, or social network potentially share *information* that yields commonality in choices (Schneider et al., 1997; Teske et al., 2007).

The foregoing examples suggest that school choices will be correlated within identifiable groups, yet few studies look directly at similarity in discrete school choices. The extent to which choices are similar is important not only for our understanding of factors that shape educational choices, but also because it affects disparities in the schools and peer groups students experience. In this paper, we introduce a new approach to describing variation in school choices: measuring within-group similarity in the *specific* schools to which students apply. This approach is useful for several reasons. First, in a system with finite options, choices are constrained by the set of available schools. A direct comparison of applications can reveal how multiple social processes manifest as specific choices. Second, families may share interest in specific schools, beyond that predicted by preferences for school characteristics. These interests may be due to salience or the influence of social networks and school personnel. Third, measuring similarity in applications

allows the researcher to identify groups with which an applicant shares specific choices. Doing so may reveal the extent to which peers such as classmates or neighbors influence school choices, and the sorting that is likely to result. Descriptive measures of similarity can help generate hypotheses about the role of shared contexts on school choices.

We use the New York City high school applications process as a case study, a context in which all 8th graders are required to rank their top choices. Nearly 80,000 apply annually to over 700 programs at 425 schools citywide, and students are centrally matched using an algorithm that incentivizes students to rank programs truthfully (Abdulkadiroğlu, Pathak, & Roth, 2005). The universality of choice and the large number and diversity of schools, neighborhoods, and populations make NYC an apt setting. The transition from middle to high school is of particular interest as students have more agency over school choices at this age and may be more influenced by school or neighborhood peers (Sattin-Bajaj, 2014; Condliffe et al., 2015).

Despite multiple reasons to expect within-group similarity in school choices, we find high school applications are surprisingly *dissimilar* within middle schools and neighborhoods in NYC. The average student's first choice high school program appears among the top three choices of 13% of their peers in the same school, and only 10% of applicants from the same neighborhood, defined by zip code. Submitting *identical* top three choices is exceedingly rare: the average student shares top three choices with less than 1 percent of their same-school peers. While we cannot identify friendship networks in our data, these results do not provide strong evidence of coordination in student applications.

We also observe variation in choice similarity between groups of students. White and Asian students are more likely to have choices in common than Black and Hispanic students, as are students whose families speak specific languages at home, such as Chinese and Russian.

High-achieving students from all racial/ethnic backgrounds are more likely to share choices with other high-achieving students in their school than lower-achieving students are to share choices. Some of these differences are expected due to residential segregation, the spatial location of schools, and the use of academic screening at some high schools. However, regressions that condition simultaneously on student characteristics, prior achievement, and location continue to find differences in choice similarity by race, home language, and achievement.

A direct implication of these findings is that students who make dissimilar choices are less likely to remain peers than those who make similar choices. In NYC, we show that White and Asian students are more likely to experience continuity in peers as they transition from middle to high school than are Black and Hispanic students. This is partially explained by prior achievement and its role in providing access to selective schools. However, racial differences in peer continuity hold even for higher-achieving students. This finding is consistent with recent studies of student enrollment patterns in Chicago (Burdick-Will, 2017; Sirer et al., 2015) and Detroit (Lenhoff et al, forthcoming) which find students from less-advantaged neighborhoods attend a diffuse set of schools citywide while students in more-advantaged neighborhoods are likely to remain together in the same schools.

Factors Affecting Similarity in School Choices

Multiple factors affect choice similarity within groups, where "groups" can be defined in multiple ways (e.g., schools, neighborhoods, demographics, prior achievement). In this section, we describe the current academic literature on these factors, and consider how they apply in the NYC high school choice context.

On the supply side, the number and characteristics of nearby schools can widen or narrow

the set of schools families are willing and able to consider. In NYC—like other large districts—the quantity and quality of options are unevenly distributed (Edwards, forthcoming). With respect to quantity, the boroughs of Manhattan and the Bronx have largely replaced comprehensive high schools with many smaller themed high schools (Jennings, 2010; Kemple, 2015). Queens and Brooklyn, in contrast, offer a mix of both small and large high schools. Some neighborhoods in the outer boroughs, including Staten Island, are geographically isolated, limiting the number of proximate choices. With respect to quality, high-performing schools are also unevenly distributed across NYC, forcing a tradeoff for some between commuting time and school quality or safety (Burdick-Will, 2017; Corcoran, 2018; Hastings, Kane, & Staiger, 2006).

On the demand side, families vary in their preferences and constraints. Both qualitative and quantitative studies have found systematic variation in revealed preferences by family background, with high-SES parents and parents of higher-achieving students more likely to apply to schools with high test scores than low-SES parents and parents of lower-achieving students (Abdulkadiroğlu et al., 2020; Bell, 2009; Glazerman & Dotter, 2017; Harris & Larsen, 2015; Hastings, Kane, & Staiger, 2006). Families also vary in their preferences for the racial composition of schools (Schneider & Buckley, 2002) and in their prioritization and conceptions of school safety (Hailey, 2020).

Some variation in choices may be due to families' sources and use of information (Corcoran & Jennings, 2019). Research has found that lower-income parents are more likely to rely on "formal" sources of information, such as district-provided materials and third-party resources, while higher-income parents rely more on "informal sources," including social networks and school personnel (Schneider et al., 1997; Teske et al., 2007). Several randomized interventions have found that providing information about school quality affects choices

(Corcoran et al., 2018; Hastings & Weinstein, 2008; Valant & Loeb, 2014; Valant & Weixler, 2020). These studies offer examples of how common sources of information can reduce (or in some cases widen) gaps in application to higher-performing schools. In NYC, school counselors are ostensibly tasked with guiding students through the high school choice process, though counselors vary in their willingness to give advice about which specific schools to apply to, and how to rank them (Sattin-Bajaj et al., 2018). Beyond counselors, students learn about high school options through fairs, open houses, and a printed directory that includes each school's location, eligibility requirements, graduation rate, and course offerings. (In 2019, the directory was replaced by an online portal).

In the transition from middle to high school, peers potentially play a role in providing information or shaping preferences. Studies have found that students have greater agency in high school choice than earlier stages, and especially so for lower-income and immigrant students (Condliffe et al., 2015; Sattin-Bajaj, 2014; Teske et al., 2007). Few studies have specifically examined the role of peers in K-12 school choice, although a study from Sweden found students' decision to apply to an upper-secondary academic track was related to the decisions of their in-group peers (Rosenqvist, 2017). Another paper by Dustan (2018) found a strong causal effect of siblings—a specific kind of peer—on high school applications in Mexico City. That paper shows its effect is driven in part by information rather than household economies of scale.[2]

Finally, more than a third of NYC high school programs screen applicants using academic or other criteria, such as an audition. These admissions criteria shape students' feasible choice sets, since they affect the likelihood of admission to a school (Sartain & Barrow, 2020).

[2] Several studies have found sibling effects on college applications in the U.S. (Goodman et al, 2016); and in Chile, Croatia, and Sweden (Aguirre & Matta, 2018; Altmejd et al., forthcoming).

Students also have access to "educational option" programs, which screen only half their students; "unscreened" or "limited unscreened" programs which do not consider academic criteria, but may give priority to students based on their current school, neighborhood, or attendance at an open house; and "zoned" programs which guarantee admission to students who live within a geographic area. Until recently, some programs gave priority admission to students who live within a particular geographic area.

Taken together, there are multiple reasons to expect similarity in school applications, as well as variation in measured similarity across groups, defined by geography, school, or other shared characteristic. Ultimately, it is an empirical question how these factors combine to yield similar or dissimilar school choice applications.

Measuring Similarity

We sought a measure of within-group similarity that treats programs or schools as distinct objects. In NYC, students apply to programs rather than schools; in 2017, there were 781 programs offered in 435 schools citywide. Some schools offer more than one program, although most (70%) offer just one. Applicants rank up to 12 programs on their application.

Drawing on a literature that develops similarity measures for sequences of categorical data, we identified three desirable properties for this measure (e.g., Brzinsky-Fay, Kohler, & Luniak, 2006; Gómez-Alonso & Valls, 2008). The first—and most important—is that the measure should capture the extent to which two applications have elements in common. Second, the measure should capture the degree of similarity in *rankings* of common elements on the application, reflecting the strength of preference for each. Third, the measure should weight commonality in *higher*-ranked elements more heavily than commonality in lower-ranked

elements. The latter property is desirable in that students are more likely to be matched to one of their top choices. (In the year of our study, roughly half of all applicants matched to their first choice, and 75% were matched to one of their top three choices).

Among the indices we analyzed, our preferred measure is a simple one: the proportion of students in group *j* whose first choice is listed among the top three choices of all other students in the group. Group *j* can refer to any collection of students, including a school, neighborhood, or borough, or demographic group within each. This measure, which we refer to as *Top3,* is calculated for each individual student *i* and then averaged over students in group *j*:

$$Top3_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(\frac{1}{n_j - 1}\sum_{k\neq i} D_{ik}\right)$$

where n_j is the number of students in group *j* and D_{ik} equals one if the first-choice school on *i*'s application is the first, second, or third choice on student *k*'s application (and zero otherwise). This measure meets our above criteria, is computationally simple, and is easy to interpret. While other similarity measures for categorical data incorporate more information and involve more complex comparisons, our results are qualitatively similar for these other measures.

We report mean values of *Top3* for all students and for select subgroups, including race/ethnicity, home language and prior achievement. Similarity is calculated with respect to *own-subgroup* or *all other* students in a school, zip code, borough, or the city. The borough and city cases serve primarily as benchmarks. As group differences in similarity are partly explained by differences in location or other confounding factors, we also used regression to calculate adjusted *Top3* scores, conditioning on student characteristics and either middle school fixed effects, zip code fixed effects, or both. This procedure is explained later.

For our main results, we calculate *Top3* similarity using program codes, since these are the actual choices students identify on their applications. For comparison, we also report results using schools as the application element of interest, as it may be that applications are more similar when considering the *school* to which students apply. (This is more relevant in parts of the city where schools offer more than one program).

A special case of D_{ik}=1 arises whenever students *i* and *k* have *identical* top three choices on their application. Friends who coordinate their application would meet this condition, as would two students who acted on the same set of school recommendations from their guidance counselor. In addition to the *Top3* results, we also measure similarity using identical (top three) application choices, again reporting results using both program and school codes.

Finally, we measure the extent to which 8th grade students attend high school with others from the same middle school or neighborhood. Since most students are matched to one of their top three choices, these outcomes are strongly affected by the degree of within-middle school or neighborhood similarity in high school applications.

Data

We used de-identified student-level applications data available through a data sharing agreement between the Research Alliance for New York City Schools and the NYCDOE. Our dataset was constructed by matching applications data from 2014-15 to demographic and other student background information. We focused on the main round of admissions, when initial preferences are submitted and when most students are matched (about 92% in 2014-15). We used 2014-15 because Corcoran et al. (2018) conducted large-scale informational interventions

during the 2015-16 and 2016-17 cycles which may have focused choices on a smaller set of schools, thereby increasing similarity.

Demographic and other student data include race/ethnicity, gender, language spoken at home, residential zip code, and 7th grade test scores in English Language Arts and math. 7th grade test scores are the ones used by screened high schools, and are the most recent available at the time of application. We used an average of the two subject test scores—each standardized to mean zero and standard deviation one—as our achievement measure. If either test score was missing, we used the non-missing score. For purposes of calculating similarity with like-scoring students, we grouped students into five quintiles of achievement using the city-wide distribution.

Roughly 76,300 students submitted an application in the main round. Our analytic sample makes two key exclusions. First, we excluded 7,919 students who had the option to continue in their current school. These students, naturally, were more likely to have similar applications if they chose to remain in their school. Second, we excluded 5,778 students applying from private schools. Finally, we excluded 348 students who were attending an alternative or special education school. The final analytic sample included 62,255 applicants from 446 traditional and charter schools.

Sample means are reported in Appendix Table 1. The sample is 15% White, 27% Black, 40% Hispanic, and 17% Asian. 31% of applicants lived in Brooklyn, 29% lived in Queens, 21% lived in the Bronx, 12% lived in Manhattan, and 7% lived on Staten Island. The most common languages spoken at home were English (55%) and Spanish (25%), but a significant minority of families spoke Chinese (7%). 19% of the sample were foreign born, 49% were female, and 3% were attending a charter school.

Results

Mean application similarity

Table 1 shows within-school and within-neighborhood, borough, and city similarity for all and select subgroups of students. The first four columns report mean *Top3* similarity using program codes, while the next four columns report the same using school codes. The top row shows mean similarity for the full analytic sample. The average student's first choice program can be found among the top three choices of about 13% of students in their school (about 1 in 8). Similarity is greater within middle schools than neighborhoods, where the comparable figure is 10%. Unsurprisingly, as the geographic area widens, application similarity declines. Mean similarity falls to 5% when students are compared to others in their borough, and 1.4% when compared to others citywide.

Application similarity varies substantially by borough, in part reflecting differences in supply. Within-school similarity in choices is highest on Staten Island (24%), followed by Manhattan (17%), Brooklyn (14%), and Queens (12%). By far the lowest similarity is in the Bronx, where the average student's first-choice program was found among the top three of only 7% of their classmate's applications (about 1 in 14). Within-neighborhood similarity is 23% in Staten Island, 11% in Manhattan, Brooklyn, and Queens; and only 4% in the Bronx. It is notable that within-neighborhood application similarity in the Bronx is only slightly higher than similarity with respect to the entire borough (3%).

The remaining rows of Table 1 show within-school and within-area similarity by subgroup, where each similarity measure is with reference to others in the same group. (Appendix Table 2 reports similarity for subgroups with respect to *all* other students in their

school or area).[3] Mean differences in application similarity by race/ethnicity are especially large. White and Asian students exhibit the highest level of within-group similarity (27% and 23%, respectively), while Black and Hispanic students make much more dissimilar choices (7% and 10%). White students are 3.7 times more likely to find their first-choice program on another White student's application in their school than Black students are to find theirs on the application of another Black student in their school. That gap widens at the neighborhood level, where they are five times as likely (25% vs. 5%).

Within-group similarity is comparable for native born and immigrant students, and males and females (not shown). Similarity scores for students in the two largest home language groups, English and Spanish, were also comparable to the city average. However, students who speak minority languages at home, such as Chinese, Russian, Bengali, and Korean, have much higher within-group similarity than do students who speak English or Spanish at home. Interestingly, these students not only have applications that are more similar to own-language students in their school; their applications are also more likely to be similar to other own-language students in their borough, and the city at large.

Some of these group differences may be driven by prior achievement, which affects access to selective schools. Indeed, Table 1 shows that high-achieving students are more likely to have choices in common with other high-achieving students in their school than are lower-achieving students. The average top quintile student could find their first-choice program among the top three choices of 19% of fellow top-quintile students in their school, while the average

[3] In most cases, similarity is lower when calculating similarity with respect to *all other* students, rather than students in the same subgroup. The measures differ least for racially isolated or otherwise segregated students who tend to be in the majority in their school or neighborhood. For example, the within-school and within-neighborhood similarity measures for Black and Hispanic students are roughly the same in Table 1 and Appendix Table 2.

student in the bottom quintile could find their first choice among the top three of only 5% of other bottom-quintile students in their school. It is perhaps less obvious why bottom-quintile students would have choices in common, in contrast to top-quintile students vis-à-vis other top quintile students, so we also contrasted similarity with *all* other students in their school (Appendix Table 2). A comparably large gap exists, suggesting the difference is not an artifact of comparisons within achievement groups.[4]

We examined whether gaps in application similarity by race/ethnicity can be explained by group differences in achievement by reporting the mean *Top3* separately for each race and achievement quintile (Appendix Table 3). In that table, we measure similarity with respect to all other students in one's own racial/ethnic group, all other students in the school, and all other students in one's quintile. The White-Black and White-Hispanic gaps are roughly the same for lower and higher achieving students when measuring similarity with respect to other students in one's racial group or school. In other words, the large gaps in application similarity by race persist conditional on prior achievement. Gaps narrow with achievement when measuring similarity with respect to other students in one's quintile, although *Top3* similarity remains 2.5 times higher for White students in the top quintile than for Black students in the top quintile.

Students apply to programs, not schools, but it is possible that similarity is greater when comparing applications to schools. The next four columns of Table 1 report mean within-group similarity using school codes, rather than programs. Since there are fewer unique schools than programs, these values do tend to be higher. For example, we find the average applicant's first-choice school is among the top three choices of about 19% of other students in their middle

[4] We also calculated similarity using within-*school* achievement quintiles. These gaps narrow, likely because there are smaller differences in achievement between quintiles within schools, versus citywide. However, the gaps remain large.

school (versus 13% when using programs). However, the differences are greatest for students in parts of the city with larger schools offering multiple programs (e.g., Staten Island and Queens), and remain virtually unchanged for students in areas where programs and schools are often synonymous (e.g., Bronx and Manhattan). Overall, the qualitative differences between race, language, and achievement groups described for programs continue to hold when using schools.

Identical applications

Our similarity measure asks whether a student's first choice program or school appears anywhere among the top three choices of another student. We also calculated the percent of students in their school, geographic area, or subgroup who submitted *identical* top three choices. Means analogous to those in Table 1 are reported in Table 2, again for programs and schools.[5]

It is rare for students in the same middle school to submit identical top three choices, and exceedingly so for certain subgroups. For the full analytic sample, we find the average student shared her top three ranked program choices with only 0.7% of her same-school peers. With 134 applicants in the average school in our sample, this translates to about 1 other student with identical choices. Consistent with Table 1, the mean proportion of students with identical applications is highest for White students, speakers of minority languages, higher-achieving students, and Staten Island residents. The proportion is lowest for Black students, low-achieving students, and Bronx residents. For example, the average Black student shared her top three ranked programs with only 0.1% of Black students in her school. The proportion of identical applications is larger when using school versus program codes but remains small. For example,

[5] Similarity with reference to the borough and city are suppressed in this table, as these values are very small.

the average student shared her top three ranked *schools* with only 2.4% of her same-school peers, or about 3 students in the average cohort size of 134.

As in the previous section, we examined whether gaps in the proportion of students with identical applications by race/ethnicity could be explained by prior achievement (Appendix Table 4). The relationship between achievement and identical applications is nonlinear, with the proportion of identical applications lowest for bottom and top-quintile students. The overall proportion is low for all achievement levels, however, and gaps between racial groups remain large.

Regression-adjusted differences in similarity

Group differences in application similarity are driven in part by residential segregation and the effective supply of schools. For example, Black and Hispanic students are over-represented in the Bronx, where the number of proximate schools is large, while White and Asian students are over-represented in Staten Island and Queens, where there are comparably fewer schools. In NYC, there is also an extensive middle school choice process in which many students opt to attend middle school outside their neighborhood. Extant sorting across middle schools could also affect subgroup differences in choice similarity. Finally, as noted earlier, many high school programs in NYC have selective admissions, which narrows the choice set for some and creates opportunities for others. Given the correlation between prior achievement, student race/ethnicity, and other indicators of family background, selective school options may also explain differences in choice similarity.

We used regression to adjust for the effects of residential location and prior achievement on within-school application similarity. These regressions include as covariates student

demographics (gender, race/ethnicity, immigrant status, language spoken at home), a cubic function of prior achievement, and (alternately) fixed effects for borough, zip code, and middle school. We also estimated a regression with both zip code and middle school effects. For each regression we used predictive margins to contrast differences in within-group similarity after accounting for the effects of observed student characteristics, middle school, and residential location. For this analysis, we measured similarity using program codes, but estimated the same regressions using schools in place of programs. (See Appendix Figures 2-4; full regression results are reported in Appendix Tables 5-10).

Predictive margins by race/ethnicity and achievement levels are plotted in Figures 1-2, with 95% confidence intervals shown; a similar plot for home language is provided in Appendix Figure 1. In each case, similarity is measured with respect to own group. For plotting the achievement margins, we chose *z*-score values of -1.3, -0.5, 0, 0.5, and 1.3, which correspond roughly to means within each quintile. Five values are shown for each subgroup. The first is the unadjusted mean similarity measure from Table 1. The second comes from a regression with covariates and borough dummies, and the third through fifth come from regressions with covariates plus zip code fixed effects, middle school fixed effects, or both. A comparison of point estimates shows the extent to which group differences in similarity are attenuated by controls for location and other student characteristics.

Although gaps in within-group similarity are reduced by the regressions, we continue to find significant differences by race/ethnicity, prior achievement, and home language. Controlling for neighborhood and middle school shrinks the gap between racial/ethnic groups (Figure 1), but even with the full set of controls, within-group and middle school similarity for White and Asian

students is 58 to 77% greater than for Black students. Within-group similarity for Hispanic students is only marginally higher than that for Black students.

Regression adjustments have a smaller effect on similarity gaps by prior achievement and language spoken at home, and high-achieving students continue to stand apart. Within-school similarity for high achieving students (1.3 standard deviations above city average) remains roughly twice that of students scoring at the city average and below (Figure 2). Students who speak a minority language at home still exhibit greater within-group similarity in school choices than do students who speak English and Spanish (Appendix Figure 1).

While regressions with middle school and/or zip code fixed effects are useful in that they permit comparisons of students with similar nearby choice options, it is important to note that they absorb some common influences that lead some students to have more similar choices than others. A regression with middle school fixed effects would remove these cross-school differences and rely only on within-school differences in similarity between groups.

Middle-to-high school peer continuity

Students who make dissimilar choices are less likely to remain peers in high school than those who make similar choices. In Table 3, we report two statistics for students in our analytic sample, overall and by subgroup: (1) the mean proportion of students in their 9th grade class who attended their middle school or resided in their zip code; and (2) the proportion who were the *only* student in their 9th grade class from their middle school or zip code.

Racial gaps in application similarity carry forward to peer continuity: White students are more likely to attend high school with other applicants from their middle school, while Black and Hispanic students are less likely to do so. For example, 1 in 5 of the average White student's 9th

grade class (19%) attended their middle school, and 1 in 6 resided in their zip code (16.5%). Only 1 in 13 (7.3%) of the average Black student's 9th grade class attended their middle school, while about 1 in 11 (9.4%) resided in their zip code. Fully 28% of Black students were the *only* student from their middle school in their 9th grade class, as compared to 7.3% for White students. 13% of Black students were the only student from their zip code in their 9th grade class, versus 7.7% of White students.

Consistent with Tables 1-2, higher-achieving students attended high schools where a greater share of their peers attended their middle school or resided in their zip code, and they were substantially less likely than lower-achieving students to be the only student from their middle school. We examined whether the differences in peer isolation by race can be explained by group differences in achievement (Appendix Figures 5-6). Higher-achieving students of all racial groups are less likely than lower-achieving students to be the only student from their middle school in 9th grade. Gaps between racial groups remains large, however, even for the highest achieving students. Interestingly, the pattern is different for neighborhoods: top quintile Black and Hispanic students were *more* likely to be the only student from their neighborhood in their 9th grade class than were other Black and Hispanic students.

Table 3 was generated using our analytic sample, which excludes students with the option to remain in the same school for 9th grade. This omits students who may be more likely to attend 9th grade with peers from their middle school or neighborhood. We repeated this analysis using the full sample of applicants (see Appendix Table 11); the results are very similar.

Discussion

This descriptive analysis departs from existing research on school choice by describing similarity within and between groups in the choice of *specific* schools. It is, to our knowledge, the first to examine school choices in this way. While related, our approach shifts the focus from correlated preferences for school characteristics toward specific schools identified as students' top choices. With this approach, one can observe whether—and to what extent—a student shares choices with others in a particular reference group. Similarity in choice applications also has direct implications for who is likely to remain peers in the transition from middle to high school.

Despite many theoretical reasons to expect similarity in school applications within middle schools, neighborhoods, and subgroups, we find high school application similarity in NYC to be surprisingly low. Within-group similarity is higher for White and Asian students (versus Black and Hispanic), for higher-achieving students, and for some minority language groups. These differences remain sizable even after controlling for factors such as residential location and prior achievement that shape choice sets. While small sets of students may coordinate top three choices with friends or classmates, our results fail to provide strong evidence of peer influence. This may be surprising to those who expect adolescents to prioritize friends when choosing schools. That said, our results are consistent with a recent survey that found few 8th graders in Chicago rated friends' plans as "very important" when choosing schools (Sartain & Barrow, 2020).

A direct consequence of these findings is that White and Asian students, and higher-achieving students of all racial/ethnic groups, are more likely to remain classmates in 9th grade than are Black and Hispanic students, and lower-achieving students. This finding mirrors Burdick-Will (2017) and Sirer et al. (2015) which found similar patterns in Chicago. There,

students from less advantaged neighborhoods traveled further to a wider variety of high schools that were often no higher performing (but in some cases were safer) than those in their own neighborhood.

Group differences in school choice similarity are not *a priori* "good" or "bad," and may simply reflect unobserved heterogeneity in underlying preferences and the supply of available schools. In some cases, families may purposefully seek out schools that provide a different peer group than the one in their middle school or neighborhood. More research is needed on whether similarity in school applications can be explained by traditional demand and supply side factors, or whether other influences are at play.

We see several opportunities for future research using measures of similarity in school choice applications. First, these measures might be combined with better data on classroom assignments and social networks to identify where and how peer effects are important. Second, these measures could be used to test for effects of informational interventions or school inputs—such as guidance counselors or teachers—on specific choices. Third, these measures might be used to identify school and neighborhood level factors that lead to commonality in choices. More generally, conditional on demand and supply-side factors, greater dissimilarity in choices could be an indicator of a school in greater need of support.

Finally, more research is needed on whether social fragmentation or peer continuity is important for adolescents' social and educational outcomes. One recent study (Felmlee et al., 2018) suggests that the disruption of social networks during school transitions can lead to social isolation and academic difficulties. Another (Stein, Burdick-Will, & Grigg, 2021) finds high school students are less likely to make early school transfers when they attend school with more of their 8th grade peers. Other studies, however, suggest that attending high school without

middle school peers allows for “fresh starts” for students who were socially isolated in middle school (Weiss & Bearman, 2007). In any case, in large urban school districts with school choice, disruptions in peer networks fall disproportionately on Black and Hispanic students, and on lower-achieving students from all racial/ethnic backgrounds.

References

Abdulkadiroğlu, A., Pathak, P., & Roth, A.E. (2005). The New York City high school match. *The American Economic Review*, *95*(2): 364-367.

Abdulkadiroğlu, A., Pathak, P.A., Schellenberg, J., & Walters, C.R. (2020). Do parents value school effectiveness? *American Economic Review,* 110(5): 1502-1539.

Aguirre, J. & Matta, J. (2018). Walking in your footsteps: Sibling spillovers in higher education choices. https://ssrn.com/abstract=3255875

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., & Smith, J. (forthcoming). O brother, where start thou? Sibling spillovers on college and major choice in four countries. *The Quarterly Journal of Economics*.

Bell, C. (2009). All choices created equal? The role of choice sets in the selection of schools." *Peabody Journal of Education*. *84*(2): 191-208.

Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal*, 6(4), 435-460.

Burdick-Will, J. (2017). Neighbors, but not classmates: Neighborhood disadvantage and the heterogeneity of educational experiences. *American Journal of Education*, *124*(1), 37-65.

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

Corcoran, S.P. (2018). School choice and commuting. Washington, D.C.: The Urban Institute.

Corcoran, S. P., & Jennings, J. L. (2019). Information and school choice. In M. Berends, A. Primus, & M. G. Springer (Eds.), *Handbook of Research on School Choice, 2nd Edition.* New York: Routledge.

Corcoran, S.P., Jennings, J.L., Cohodes, S.R., & Sattin-Bajaj, C. (2018). Leveling the playing field for high school choice: Results from a field experiment of informational interventions. *NBER Working Paper No. 24471*.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics* 134: 372-391.

Edwards, D. S. (forthcoming). Just out of reach? Unrestrained supply, constrained demand, and access to effective schools in and around Detroit. *Educational Evaluation and Policy Analysis*.

Felmlee, D., et al. (2018). Falling behind: Lingering costs of the high school transition for youth friendships and grades. *Sociology of Education*. *91*(2): 159-182.

Glazerman, S. and D. Dotter (2017). Market signals: Evidence on the determinants and consequences of school choice from a citywide lottery. *Educational Evaluation and Policy Analysis 39*(4): 593—619.

Gómez-Alonso C., Valls A. (2008). A similarity measure for sequences of categorical data based on the ordering of common elements. In: Torra, V. & Narukawa Y. (eds) *Modeling decisions for artificial intelligence*. Berlin: Springer.

Goodman, J., Hurwitz, M., Smith, J., & Fox, J. (2016). The relationship between siblings' college choices: Evidence from one million SAT-taking families. *Economics of Education Review*, *51*, 125–135.

Hailey, C.A. (2020). Racialized prisms: A randomized experiment on families' perceptions of schools. Working paper, University of Texas at Austin.

Harris, D. N., & Larsen, M. F. (2015). What schools do families want (and why)? *Education Research Alliance for New Orleans*, 66.

Hastings, J.S., Kane, T.J., & Staiger, D.O. (2006). Parental preferences and school competition: Evidence from a public school choice program. Unpublished working paper.

Hastings, J.S., & Weinstein, J.M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, *123*(4), 1373–1414.

Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227-247.

Kemple, J. J. (2015). High school closures in New York City: Impacts on students' academic outcomes, attendance, and mobility. New York: Research Alliance for New York City Schools.

Lenhoff, S.W., Singer, J., Pogodzinski, B. & Cook, W. (forthcoming) Exiting Detroit for school: Inequitable choice sets and school quality. *Journal of Education Policy*.

Rosenqvist, E. (2018). Two functions of peer influence on upper-secondary education application behavior. *Sociology of Education 91*(1), 72-89.

Sartain, L. & Barrow, L. (2020). The pathway to enrolling in a high-performance high school: understanding barriers to access. Federal Reserve Bank of Chicago WP 2020-32.

Sattin-Bajaj, C. (2014). *Unaccompanied minors immigrant youth, school choice, and the pursuit of equity*. Harvard Education Press.

Sattin-Bajaj, C., Jennings, J.L., Corcoran, S.P., Baker-Smith, E.C. & Hailey, C. (2018). Surviving at the street level: How counselors' implementation of school choice policy shapes students' high school destinations. *Sociology of Education* 91(1): 46-71.

Schneider, M. & Buckley, J. (2002). What do parents want from schools? Evidence from the internet. *Educational Evaluation and Policy Analysis* 24(2): 133-44.

Schneider, M., Teske, P., Roch, C., & Marschall, M. (1997). Networks to nowhere: Segregation and stratification in networks of information about schools. *American Journal of Political Science. 41*(4), 1201–1223.

Sirer, M.I., et al. (2015). The currents beneath the 'rising tide' of school choice: An analysis of student enrollment flows in the Chicago public schools. *Journal of Policy Analysis and Management. 34*(2), 358-377.

Stein, M. L., Burdick-Will, J., & Grigg, J. (2021). A choice too far: Transit difficulty and early high school transfer. *Educational Researcher*, *50*(3), 137-144.

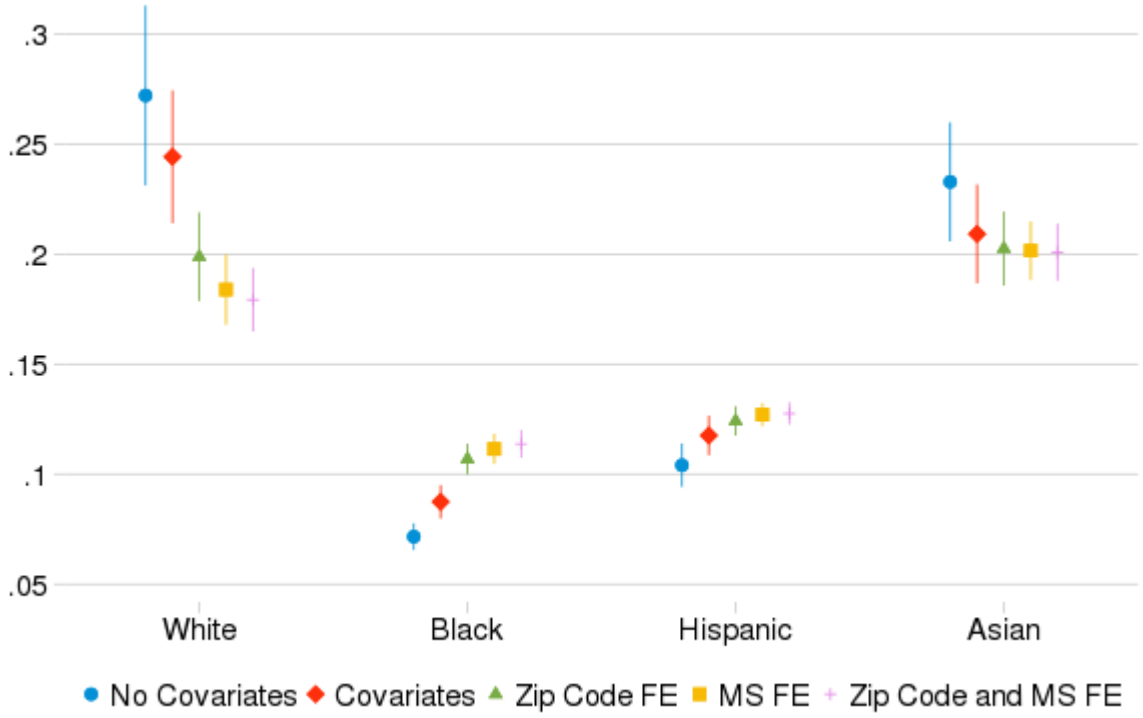
Teske, P., Fitzpatrick, J., & Kaplan, G. (2007). *Opening doors: How low-income parents search for the right school*. Seattle: Center on Reinventing Public Education.

Valant, J. & Loeb, S. (2014). Information, choice, and decision-making: Field experiments with adult and student school choosers. Unpublished working paper.

Valant, J. & Weixler, L. (2020). Informing school-choosing families about their options: A field experiment from New Orleans. EdWorkingPaper: 20-325.

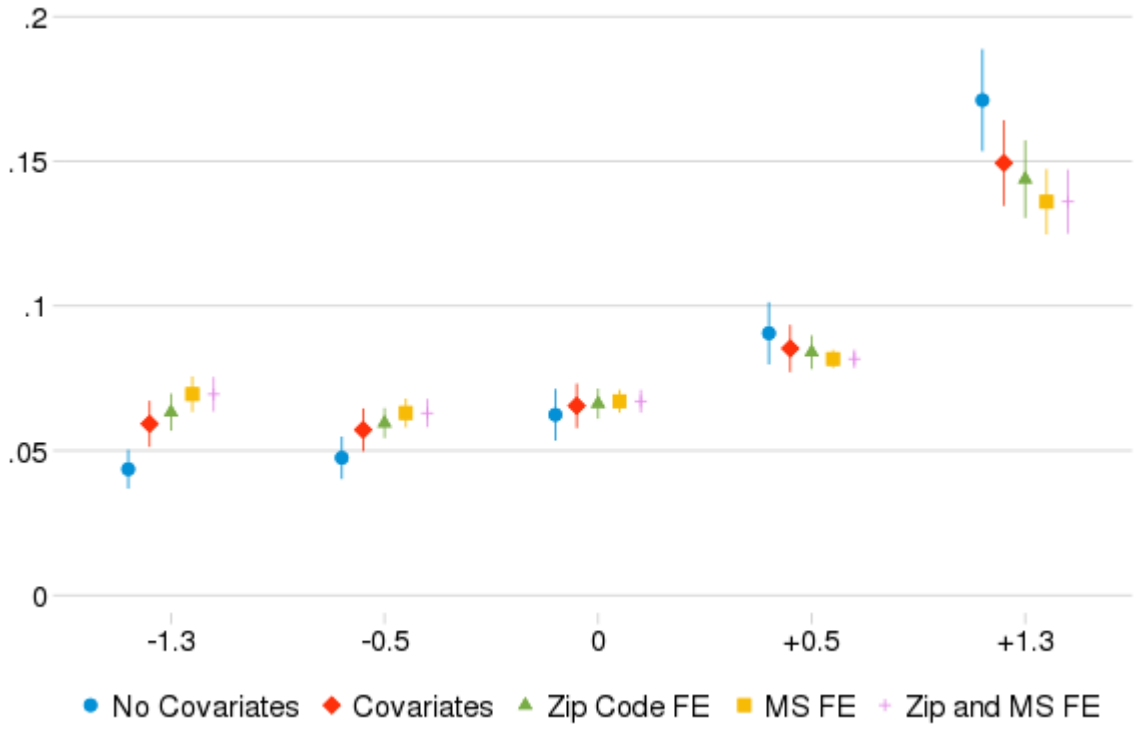
Weiss, C., & Bearman, P. (2007). Fresh starts: Reinvestigating the effects of the transition to high school on student outcomes. *American Journal of Education, 113*(3), 395-421.

Figure 1: Within-School Similarity by Race/Ethnicity



Note: *Top3* scores are calculated using the population of students applying to high school from public middle schools and charter schools, but the regressions are estimated using observations in the analytic sample, which excludes students in schools with a continuing 9th grade and those with missing data. Plotted points are: (1) the unadjusted mean similarity measure from Table 1; (2) predicted similarity scores from a regression with borough dummies and the following covariates: sex, race/ethnicity, immigration status, special education, language spoken at home, and a cubic function of average 7th grade reading and math test scores; (3-5) predicted similarity from a regression with the same covariates and either zip code fixed effects, middle school fixed effects, or both. 95% confidence intervals shown, accounting for clustering at the middle school level.

Figure 2: Within-School Similarity by Achievement Z-Score



Note: see notes to Figure 1 for details.

Table 1: Mean High School Application Similarity, 2014-15

		Programs: with respect to *all* students in the same:				Schools: with respect to *all* students in the same:				
		School	Zip code	Borough	City	School	Zip code	Borough	City	N
	All	0.128	0.102	0.048	0.014	0.191	0.162	0.081	0.022	62,255
MS borough	Manhattan	0.168	0.106	0.059	0.015	0.171	0.110	0.061	0.016	19,435
	Brooklyn	0.141	0.113	0.047	0.017	0.200	0.168	0.076	0.027	7,319
	Queens	0.117	0.105	0.046	0.016	0.201	0.185	0.090	0.029	18,273
	Staten Island	0.235	0.228	0.105	0.008	0.515	0.506	0.264	0.019	4,091
	Bronx	0.067	0.038	0.027	0.009	0.075	0.043	0.030	0.010	13,137
		Programs: with respect to *own group* in the same:				Schools: with respect to *own group* in the same:				
Subgroup:		School	Zip code	Borough	City	School	Zip code	Borough	City	N
Race/ Ethnicity	White	0.272	0.248	0.137	0.039	0.394	0.370	0.205	0.057	9,185
	Black	0.072	0.050	0.034	0.012	0.117	0.090	0.065	0.021	16,595
	Hispanic	0.104	0.075	0.037	0.011	0.147	0.114	0.059	0.016	25,200
	Asian	0.233	0.203	0.149	0.059	0.330	0.298	0.212	0.084	10,382
Home Language	English	0.120	0.097	0.046	0.013	0.181	0.155	0.079	0.020	34,135
	Spanish	0.113	0.079	0.039	0.012	0.157	0.119	0.062	0.017	15,837
	Chinese	0.288	0.249	0.219	0.090	0.401	0.361	0.314	0.129	4,080
	Russian	0.326	0.292	0.234	0.132	0.434	0.406	0.319	0.171	977
	Bengali	0.208	0.175	0.128	0.070	0.256	0.217	0.154	0.081	1,531
	Korean	0.298	0.299	0.228	0.179	0.461	0.455	0.356	0.288	366
	Creole	0.107	0.090	0.061	0.040	0.206	0.197	0.157	0.103	449
Ach. Quintile (citywide)	Lowest	0.047	0.033	0.011	0.003	0.077	0.061	0.023	0.005	12,005
	Second	0.051	0.038	0.013	0.004	0.085	0.069	0.028	0.007	12,011
	Third	0.063	0.048	0.019	0.005	0.105	0.089	0.038	0.009	11,740
	Fourth	0.077	0.062	0.029	0.008	0.130	0.112	0.054	0.014	11,561

Highest	0.189	0.176	0.122	0.045	0.235	0.220	0.149	0.051	11,713

Notes: Authors' calculations using applications data provided by the Research Alliance for New York City Schools. Sample includes all 8th grade students attending a NYC public middle school, including charter schools, who applied in the main round of admissions and did not have the option to remain in their school. Reported values are averages of the *Top3* score for a population of students (listed in the leftmost column) with respect to other students in a school or geographic area (shown at top). The first four columns use applications to high school programs; the next four columns use high school codes rather than programs.

Table 2: Identical Top Three Program and School Choices, 2014-15

		Programs: with respect to *all* students in the same:		Schools: with respect to *all* students in the same:	
		School	Zip code	School	Zip code
	All	0.007	0.006	0.024	0.022
MS borough	Manhattan	0.004	0.001	0.005	0.001
	Brooklyn	0.006	0.004	0.014	0.011
	Queens	0.003	0.003	0.013	0.013
	Staten Island	0.058	0.058	0.226	0.225
	Bronx	0.002	<0.001	0.002	0.001
		Programs: with respect to *own group* in the same:		Schools: with respect to *own group* in the same:	
Subgroup:		School	Zip code	School	Zip code
Race/ Ethnicity	White	0.033	0.031	0.101	0.098
	Black	0.002	0.001	0.005	0.003
	Hispanic	0.004	0.003	0.013	0.010
	Asian	0.008	0.007	0.028	0.025
Home Language	English	0.009	0.008	0.029	0.028
	Spanish	0.005	0.002	0.014	0.009
	Chinese	0.011	0.008	0.035	0.029
	Russian	0.021	0.017	0.064	0.057
	Bengali	0.006	0.003	0.010	0.009
	Korean	0.015	0.022	0.071	0.078
	Creole	0.003	0.001	0.006	0.004
Ach. Quintile (citywide)	Lowest	0.008	0.006	0.017	0.016
	Second	0.007	0.006	0.018	0.016
	Third	0.012	0.010	0.026	0.024
	Fourth	0.012	0.011	0.033	0.032
	Highest	0.014	0.012	0.043	0.041

Notes: Authors' calculations using applications data provided by the Research Alliance for New York City Schools. For sample definition and sample sizes, see Table 1. Reported values are the mean percent of students who share identical top three choices in a school or geographic area. The first four columns use applications to high school programs; the next four columns use school codes rather than programs.

Table 3: Middle School to High School Peer Continuity

	% of Gr9 from MS	Only student from MS	% of Gr9 from zip code	Only student from zip code
All students	0.102	0.183	0.110	0.102
Brooklyn	0.082	0.192	0.107	0.095
Manhattan	0.082	0.255	0.066	0.207
Queens	0.112	0.094	0.103	0.085
Staten Island	0.267	0.026	0.261	0.025
Bronx	0.079	0.301	0.103	0.102
White	0.190	0.073	0.165	0.077
Black	0.073	0.275	0.094	0.131
Hispanic	0.094	0.204	0.109	0.105
Asian	0.090	0.079	0.091	0.071
English	0.112	0.212	0.114	0.120
Spanish	0.090	0.186	0.110	0.090
Chinese	0.079	0.056	0.089	0.064
Russian	0.115	0.050	0.121	0.047
Bengali	0.073	0.153	0.072	0.091
Korean	0.121	0.027	0.106	0.052
Creole	0.055	0.140	0.099	0.065
Citywide quintile:				
Lowest	0.086	0.227	0.115	0.105
Second	0.086	0.217	0.109	0.108
Third	0.099	0.180	0.111	0.094
Fourth	0.117	0.161	0.116	0.097
Highest	0.119	0.109	0.093	0.092

Notes: Authors' calculations using applications data provided by the Research Alliance for New York City Schools. For sample definition and sample sizes, see Table 1. Reported values are: (1) the mean proportion of students in the students' enrolled 9th grade school who attended the same middle school or resided in the same zip code as the applicant; and (2) the proportion of students who were the *only* student from their middle school or zip code in their enrolled 9th grade school.