The Racial Gap in Friendships Among High-Achieving Students

Weonhyeok Chung and Jeonghyeok Kim*

September 25, 2024

Abstract

High-achieving minority students have fewer friends than their majority counterparts. Exploring patterns of friendship formation in the Add Health data, we find strong racial homophily in friendship formations as well as strong achievement homophily within race. However, we find that achievement matters less in cross-racial friendships. As a result, high-achieving Black students lose Black friends as they move away from the mean achievement of their group, but do not gain high-achieving White friends in offsetting fashion. We find that high-achieving Black students have fewer friends, mainly due to the fact that they are exposed to fewer high-achieving peers within their own race. Using existing causal returns to friendship in the literature, we find that this could account for as much as 5–9 percentage points (16-33%) of the racial earnings gap observed among high achievers.
JEL: I21, J15
Keywords: friendship formation, homophily, racial friendship gap, racial earnings gap

*Weonhyeok Chung: Korea Institute for International Economic Policy, Sejong, South Korea (weonhyeok.chung@gmail.com); Jeonghyeok Kim: Department of Economics, University of Houston, Houston, Texas, USA (jkim124@uh.edu). We thank Chinhui Juhn, Yona Rubinstein, Nathan Canen, Vikram Maheshri, Willa Friedman, Leandro Carvalho, Carolina Arteaga, Ko Sugiura, Osea Giuntella, Fan Wang, Su Hwan Chung, Andrea Szabo, Radek Paluszynski, Anton Badev, Richard Murphy, Cody Tuttle, Joan Llull, and seminar participants at the University of Houston, Southern Economic Association 92nd Annual Meeting (2022), Korea Labor Institute and Korea Institute for International Economic Policy, and European Winter Meeting of the Econometric Society (2023) for valuable comments. This paper previously circulated as "Friendship Formation by Race and Abilities."

1 Introduction

Having friends in school is of paramount importance for well-being of adolescents (Anelli and Peri 2019; Badev 2021; Bifulco, Fletcher, and Ross 2011; Calvó-Armengol, Patacchini, and Zenou 2009; Carrell, Hoekstra, and Kuka 2018; Imberman, Kugler, and Sacerdote 2012; Nakajima 2007). Recent works by Chetty et al. (2022a, 2022b) even find that friendships play a crucial role in shaping economic mobility and inequality, emphasizing that connecting with peers from higher socioeconomic backgrounds increases the likelihood of upward mobility. However, segregation based on socioeconomic status and race limits minority children's access to peers from higher socioeconomic backgrounds (Ananat 2011; Angrist and Lang 2004; Cutler and Glaeser 1997). Even in areas with less segregation, a "friending bias" persists, wherein individuals tend to form friendships with others from the same race and socioeconomic class (Mayer and Puller 2008; McPherson, Smith-Lovin, and Cook 2001; Mele 2020). These dynamics contribute to the perpetuation of socioeconomic disparities despite efforts at desegregation.

The disparity extends beyond the traits of friendships (Currarini, Jackson, and Pin 2010; Marsden 1987). As Figure 1 shows, Black students in the Add Health data have fewer friends than White students, and the gap grows with achievement. Some attribute this pattern to evidence that high-achieving minority students may face social ostracization from their peers, a phenomenon known as "acting White" (Fryer Jr and Torelli 2010). Considering the importance of the number of friendships (Lleras-Muney et al. 2020), the disadvantage minority students encounter is further exacerbated, alongside differences in the characteristics of the friendships they form, even before they enter the labor market. Consequently, comprehending the factors influencing friendship formation and the roots of disparities in friendships among different racial groups is crucial in gaining insight into the well-being of adolescents, economic mobility, and inequality.

Using social-network data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), we study the racial gap in the number of friendships, particularly among high-achievers, and its implications for later labor market outcomes. We first construct a framework of friendship formation taking into account academic achievement and race. To do

this, we use dyadic regression, where we construct a dataset of all possible pairs of students within the school. Specifically, we investigate how differences in academic achievement interact with the race of students in the formation of friendships. Based on the framework, we explore the reasons for the racial gap in friendships and the types of friends they form. We also find implications for the racial earnings gap.

Our results show strong patterns of racial homophily and homophily in academic achievement, in line with previous studies (Currarini, Jackson, and Pin 2009; Flashman 2012a; Mele 2020; Smirnov and Thurner 2017). The probabilities of friendship formation are 1.11% for Black-Black pairs and 0.86% for White-White pairs among all possible pairs in school. In contrast, the probability of friendship formation is lower, at 0.24%, for Black-White pairs. An increase of one unit in the GPA difference within pairs of students corresponds to a 0.164 percentage point decrease in the likelihood of friendship formation within the school when the overall mean is 0.694%. Notably, the relationship between GPA disparity and friendship formation is more prominent among individuals of the same race than those of different racial backgrounds, with both Black-Black pairs and White-White pairs showing similar sensitivity to GPA differences. As a result, high-achieving Black students lose Black friends as they move away from the mean achievement of their group, but do not gain high-achieving White friends in offsetting fashion. These patterns of friendship formation suggest that racial homophily, together with lower rates of exposure to peers of their own race, is one factor explaining why high-achieving minority populations have fewer friends compared to their majority counterparts.

Based on the estimates from the dyadic regression and the student composition in each school, we conduct a counterfactual analysis to investigate how the racial gap in friendships changes for high-achieving students as the student composition changes relative to low-achieving students. We first equalize the proportion of Black and White students in school (level effects), and then equalize the GPA distribution of Black students with that of White students (composition effect). We find that level effects are dominant in explaining the friendship gap since equalizing proportion significantly changes student racial composition in school. Through our counterfactual analyses incorporating both level and composition effects, we observe changes in the friendship gap (White minus Black) of high-achieving students from 1.05 to 0.07 for high-achieving friends

and from -0.26 to -0.15 for low-achieving friends. Thus, if the GPA distribution and the number of Black and White students were equal, the gap in friends for high-achieving students would decrease by 0.87 in total.

Lastly, we assess the extent to which the Black-White earnings gap can be attributed to the difference in the number of friends. Our analysis includes a sample primarily comprising individuals in their early twenties with an average annual income of $37,000. In our sample, Black workers exhibit 28.7% lower yearly earnings compared to White workers. Following Lleras-Muney et al. (2020), we estimate the causal return to one more friend using the increasing probability of forming friendships at similar ages as an instrumental variable. Our estimates for both Black and White individuals fall within the bounds of the causal returns to friendship, ranging from 6.50% to 13.67%, as calculated by Lleras-Muney et al. (2020). Given our estimated reduction in the expected friendship gap in counterfactual scenarios (0.87 friends) and the causal impact of one more friend on earnings, the wage gap is expected to decrease in our counterfactual scenario from 28.7% to 19.3%-24.2%, representing a reduction of the gap by 4.5-9.4 percentage points (15.7%-32.8%).

In this paper, we contribute to three strands of the literature. First, we contribute to the literature on disparities in friendships among demographic groups, especially focusing on the role of homophily in social networks. Extensive research has examined homophily across various dimensions, such as race and socioeconomic status (Chetty et al. 2022b; Marmaros and Sacerdote 2006; Mayer and Puller 2008; McPherson, Smith-Lovin, and Cook 2001; Mele 2020). Moreover, Currarini, Jackson, and Pin (2009) highlight that minority groups, in general, tend to have fewer connections. Building upon this, our study presents how homophily based on race and academic achievement, when coupled with the distribution of majority and minority students, results in fewer friendships for high-achieving minorities.

Our paper also contributes to the literature on the influence of peer pressure on youths' behavior (Bursztyn, Egorov, and Jensen 2019; Bursztyn and Jensen 2015; Coleman 1961). Particularly, Austen-Smith and Fryer Jr (2005) and Eguia (2017) theorize that high-performing Black students induce peer-group rejection (the acting-White hypothesis). Fryer Jr and Torelli (2010) supports the hypothesis with empirical evidence that high-achieving Black students tend to

have fewer friends compared to their White counterparts. Contrary to this perspective, Andrews and Swinton (2014) find limited evidence that Black students experience more substantial peer rejections based on academic achievement. Research in sociology also challenges the oppositional-culture framework (Tyson and Lewis 2021), finding that Black students are as achievement-oriented as their White peers (Diamond and Huguley 2014; Flashman 2012b; Hanselman et al. 2014). We contribute to the literature by demonstrating that the lower number of friends for high-achieving Black students can be explained by two factors: the racial composition of the student body and the GPA distribution of Black and White students.

We contribute to the literature on the racial earnings gap (Altonji and Blank 1999; Bayer and Charles 2018; Brown 1984; Juhn, Murphy, and Pierce 1993; Lang and Lehmann 2012; Neal and Johnson 1996), particularly due to the peers and social network (Arrow and Borzekowski 2004; Calvó-Armengol and Jackson 2004; Holzer 1987). We demonstrate that high-achieving Black students tend to have fewer friends because of the smaller proportion of them in schools, their lower average GPA, and homophily in friendship formation. Building upon the findings of Lleras-Muney et al. (2020) regarding the influence of friendship networks on earnings, we highlight the substantial role of racial difference in friendship to the racial earnings gap. To the best of our knowledge, this paper is the first to quantify the significance of the racial disparity in friendship networks in secondary schools for understanding earning disparities between racial groups.

The remainder of this paper proceeds as follows. In Section 2, we describe the dataset in detail. Section 3 describes the dyadic regression and presents empirical outcomes. In Section 4, we explore the implications for racial gaps in friendships and earnings by performing counterfactual exercises. We conclude in Section 5.

2 Data

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative sample of adolescents in grades 7-12 during the 1994/95 school year. The survey consists of in-school data and in-home survey data. We use the in-school data for our analysis of friendship formation and in-home survey data for labor market analysis. We restrict our sample

to students who reported GPA and demographic variables including race, age, and sex. To understand the Black-White friendship formation difference, we focus on non-Hispanic Black and non-Hispanic White students.[1]

In Figure 2 (a), we show the GPA distribution of Black and White students. We calculate the average GPA based on English, math, history, and science as a measure of achievement. The GPA scale ranges from 1 to 4. We find that White students constitute the majority of the sample, at 78.9%, while Black students constitute 21.1%. In Figure 2 (b), we equalize the proportion of Black and White students to facilitate a direct comparison of GPA distributions between the two racial groups. Our findings reveal that White students are more likely to have a higher GPA, while Black students tend to be more concentrated in the middle of the GPA distribution. Specifically, the average GPA of White students is 2.9, which is higher than that of Black students at 2.6.

The in-school data consists of information on the social network of each participant. The survey asks students to list up to five male and five female friends. In this study, we define friendship as a binary variable, taking the value 1 when an individual has been nominated as a friend by another within the same school.[2] On average, Black and White students have 3.04 and 3.80 friends, respectively.

To analyze the distribution of friendship formation and GPA differences at the pairwise level, we generate all possible pairs of students within each school and match their individual-level characteristics. Importantly, the pairs are directional. The friendship measure represents a friendship nomination from student i to student j. For instance, student i can nominate student j as a friend, but it does not necessarily mean that the nomination is reciprocated. From the pool of potential pairs of students within each school, we categorize them into four groups based on the racial composition: White-White, Black-Black, White-Black, and Black-White. In total, we have 141 schools, with an average of 395 Black or White students per school. Consequently, the overall number of potential pairs amounts to 36,183,256.

[1] See Table A.1 for summary statistics at the individual level.

[2] Students can list up to five male friends and five female friends. While a student, as the sender, can list up to ten friends, the friendship measure we use is based on the number of nominations received. For instance, if there are 400 students in a particular school, the maximum number of friendship nominations that one student can receive is 399, as they cannot nominate themselves.

Table 1 provides summary statistics at the pairwise level. First, it reveals that White-White pairs constitute 73.29% of all potential pairs among White or Black students. Black-Black pairs account for 9.91%, while both White-Black and Black-White pairs represent 8.40% each. The average rate of friendship formation across all potential pairs is 0.51%. This means that out of all possible pairs of students, 0.51% result in the formation of friendships. Black-Black pairs exhibit a higher friendship-formation rate at 0.78%, while White-White pairs have a slightly lower rate of 0.57%. Friendship-formation rates between students of different races are lower than those within the same racial group, with rates of 0.10% for Black-White pairs and 0.13% for White-Black pairs. Last, there is no significant difference between groups in terms of GPA differences between pairs. The average GPA difference for all possible Black-Black pairs is 0.8, whereas for the other types of pairs, it is 0.9.

We further restrict our sample for labor market analysis to include individuals who report measures of cognitive skills, social skills, education level, and earnings. Cognitive skills are assessed using the Add Health Picture Vocabulary Test (AHPVT) score, while social skills are measured through self-reported extroversion levels. Earnings are measured as total earnings from wages or salary in the year prior to the survey, for respondents aged 18-26.[3]

3 Friendship Formation

3.1 Empirical Strategy

We focus on analyzing patterns in the formation of friendships, with a specific emphasis on the role of race, achievement, and the interaction between those two. In our analysis, we treat each potential dyad of students within each school as an individual observation. This dyadic approach enables us to examine how the probability of forming friendships differs across racial pairs and how it changes as the difference in achievement increases between the two students.

In Figure 3, we illustrate how the proportion of dyads forming friendships varies by the difference in achievement and by race pair (Black-Black, White-White, or Black-White/White-Black).[4] A few patterns are worth noting. First, same-race pairs are more likely to form

[3] See Appendix A.7 for details on the variable descriptions used in the labor market analysis.
[4] Appendix A.2 provides a detailed explanation of our methodology for calculating the likelihood of friendship

friendships. For instance, approximately 0.9% of dyads consisting of both Black students with identical GPAs become friends. Similarly, approximately 0.75% of dyads composed of both White students with identical GPAs end up as friends. In contrast, approximately 0.2% of dyads consisting of students from different races (Black-White or White-Black) with identical GPAs end up as friends. Second, the probability of friendship formation declines as the difference in GPA increases. This declining trend is comparable between White-White and Black-Black pairs. In other words, GPA differences in friendship formation matter similarly for White-White and Black-Black pairs of students. Conversely, the difference in GPA has less significance in the formation of friendships across different races compared to the formation within the same race.

The dyadic regression equation takes the following form:

$$\begin{aligned} G_{ij} = {} & \alpha_0 + \beta_0 d_{ij} + \beta_{bb} d_{ij} \times B_i B_j + \beta_{bw} d_{ij} \times B_i W_j + \beta_{wb} d_{ij} \times W_i B_j \\ & + \alpha_{bb} B_i B_j + \alpha_{bw} B_i W_j + \alpha_{wb} W_i B_j + X_{ij}\Gamma + \lambda_s + \varepsilon_{ij} \end{aligned} \tag{1}$$

where G_{ij} is an indicator variable that takes the value of 1 if student i designates student j as a friend. The variable d_{ij} represents the difference in achievement (GPA) between students i and j expressed as absolute values. The analysis includes two racial categories: Black and White, denoted by B and W respectively. The directional links between students can be categorized into four types: B_iB_j, B_iW_j, W_iB_j, and W_iW_j. For instance, the indicator B_iW_j indicates that student i (the friendship sender) is Black and student j (the friendship receiver) is White. The omitted category in the analysis is White-White pairs of students. The vector X_{ij} incorporates a set of control variables at the dyad level, such as the difference in age and the gender pair, and individual level, including gender, age, and average GPA of senders and receivers. We also include school fixed effects, λ_s. We cluster the standard errors at the school level to account for potential dependence within schools.[5]

formation, accompanied by an illustrative example.

[5] Our regression analysis assumes that each student in a school has an equal probability of encountering and potentially forming a friendship with any other student. Under this assumption, the coefficients in the regression model represent the preference parameters associated with the characteristics of both the sender (i) and the receiver (j) as well as the disparities in their characteristics. However, one might be concerned about the assumption of equal probability. To alleviate this concern, we conduct the same analysis among students within the same extracurricular activity and find consistent results with our baseline results. See Appendix A.5 for more details.

As equation (1) excludes White pairs of students, the parameter α_0 represents the baseline probability of friendship formation for White pairs. The remaining parameters in the α group, namely α_{bb}, α_{bw}, and α_{wb}, capture the deviation in the probability of forming friendships for the respective race pairs (Black-Black, Black-White, and White-Black) compared to the baseline probability of White pairs. Regarding the β parameters, β_0 captures the extent to which differences in achievement matter for White students. If β_0 is negative, it means that White students are less likely to list other White students as friends as the difference in achievement increases. β_{bb} captures how much differences in achievement matter differently for Black students listing Black students as friends, compared to White students listing White students. Additionally, if there is racial asymmetry in how differences in achievement matter for friendship formation, the estimates of β_{bw} and β_{wb} will differ.

3.2 Results

Table 2 presents the results of the estimation of equation (1). In column (1), we find that as the difference in achievement (measured by GPA) between two students increases, the likelihood of forming friendships decreases.[6] Specifically, for each one-unit increase in GPA difference, the likelihood of friendship formation in the school decreases by 0.189 percentage points compared to the overall mean of 0.513%. The constant term represents the probability of friendship formation when there is no difference in GPA between students. In column (2), we further control for individual- and dyad-level demographic variables and school fixed effects. The coefficient remains similar. Specifically, the analysis shows that the probability of friendship formation decreases by 0.164 percentage points for each one-unit increase in GPA difference. These findings suggest a negative relationship between the difference in achievement and the likelihood of forming friendships, even after considering various controls.

In columns (3) and (4), we examine whether the difference in achievement has a varying association with friendship formation for pairs of Black students, pairs of White students, and pairs of Black and White students. We investigate this association both without additional covariates (column 3) and with them (column 4). The results indicate that regardless of whether

[6] One might be concerned about endogeneity when using GPA as a measure of achievement affecting friendship since GPA could be affected by friendship formation. To alleviate the concern, we alternatively use mother's education as a measure of achievement and find qualitatively similar results. See Appendix A.4 for more details.

the covariates are controlled for, the difference in GPA has a similar relationship to friendship formation for pairs of Black students and pairs of White students. When the GPA gap increases by 1, White-White and Black-Black pairs are 0.189 and 0.172 percentage points less likely to form friendships, respectively. On the other hand, the influence of GPA differences on friendship formation is weaker for pairs of different races compared to pairs of the same race. Specifically, Black-White and White-Black pairs are 0.031 percentage points less likely to form friendships when the GPA difference increases by 1.[7]

We find that within-race friendships are more prevalent compared to across-race friendships, consistent with the concept of racial segregation (e.g., Currarini, Jackson, and Pin 2009; Marmaros and Sacerdote 2006). The baseline probability of forming friendships within Black-Black pairs is approximately 1.112%, while that for White-White pairs is 0.862%. In contrast, the baseline probability of friendship formation across races is much lower, ranging from 0.235% to 0.259%. The findings highlight the presence of racial homophily in friendship formation, where individuals tend to form friendships more frequently with others of the same race than with individuals of different races. In column (5), we present the estimation results with individual fixed effects for both friendship senders and receivers. This accounts for the correlation between race and the possibility that certain students may be more likely to consider others as friends and some students may be more popular within the school. Despite controlling for these fixed effects, the estimated coefficients remain qualitatively similar.

One important implication of the regression results is that students might encounter different trade-offs for increasing their GPA. For instance, a Black student with the average GPA of their group (GPA 2.6) might lose friends if they increase their GPA and move away from the mean achievement of Black students, while a White student with the same GPA might not because of the difference in GPA distribution across racial groups. Using the expected number of friends from the regression results in column (4) of Table 2, we explore the possible changes in the number of friendships across the GPA distribution by race. Figure 4 presents the difference in the number of expected friends between students with the average GPA of Black students

[7] Importantly, we compare the racial groups on a percentage point level. Given that Black pairs exhibit a higher baseline probability of friendship formation, a comparable number of percentage points indicates that Black pairs are less concerned about GPA differences. However, we focus on the percentage points level comparison, as our primary interest lies in the actual number of friends formed.

and students with a GPA that is higher or lower by 1 than this average. While White students gain more expected number of friends as they increase their GPA, Black students lose Black friends. This loss is not fully compensated by the gain of White friends. In total, Black students lose 0.05 friends while White students gain 0.31 friends.[8] On the other hand, White students lose 0.25 more friends as their GPA decreases compared to Black students. While this is a descriptive analysis, it indicates that students might have different incentives for studying when it comes to popularity due to homophily and different GPA distributions by race. Moreover, it partly explains why high-achieving minority students have fewer friends than their majority counterparts.

The aforementioned findings are robust across various alternative specifications, as detailed in the appendix. We first explore heterogeneity in friendship formation across schools with different racial compositions in Appendix A.3, finding high-achieving Black students tend to have fewer friends only in Schools with a Black minority. Moreover, these alternative specifications include (i) using mother's education as a performance measure (Appendix A.4); (ii) comparing students within the same extracurricular activities (Appendix A.5); (iii) using mutual friendships instead of one-side friendship nominations (Appendix B.1); (iv) accounting for asymmetry of GPA between low-high and high-low friendship nominations (Appendix B.2); (v) using semiparametric GPA measures as low and high (Appendix B.3); and (vi) including Hispanic students in the sample (Appendix B.4). Across all these alternative specifications, the main findings regarding the relationship between GPA differences and friendship formation for race pairs remain consistent.

[8] It needs to be interpreted with caution. Since it is based on the expected number of friends from a simple model, it does not comprehensively cover all aspects of friendship formation. For instance, our model does not account for the overall popularity of high-achievers. The actual number of friends would likely increase as students achieve higher GPAs. However, as we show in the next section, our simple model effectively captures the actual racial differences in the number of friends. Therefore, our analysis is valid as long as our focus is on racial comparison.

4 The Racial Friendship Gap: Reasons and Earnings Implication

In this section, we conduct a counterfactual analysis to investigate the reasons why high-achieving Black students have fewer friends than their White counterparts. We first calculate the simulated number of friends based on the regression results from the previous section. Then, we decompose the friendship gap using racial proportions in school and the distributions of GPA for Black and White students. To account for the difference in the number of friends, we categorize friends into two groups based on their GPA: those with GPA ranging from 1 to 3 (considered low-achieving friends) and those with GPA ranging from 3 to 4 (considered high-achieving friends). Moreover, we estimate equation (1) for each school separately, using school-specific coefficients to isolate the unique characteristics of each school.[9]

We define the number of friends of student j as follows:

$$n_j = \sum_{i \in s} (G_{ij}) \tag{2}$$

where s is the school that students i and j are in. Based on the estimated parameters in equation (1), the number of Black and White students, and the GPA distributions, we compute the number of simulated friends ($\tilde{n}$) for a student with GPA of x with the following equation:

$$\begin{aligned}\tilde{n} =& n(x, \hat{\alpha}, \hat{\beta}; N_B, N_W, f_B, f_W) \\ =& \sum_{s \in E} N_B^s \times \int_{lb}^{ub} \left[(\hat{\alpha}_0^s + \hat{\alpha}_{BB}^s B + \hat{\alpha}_{WB}^s W) + (\hat{\beta}_0^s + \hat{\beta}_{BB}^s B + \hat{\beta}_{WB}^s W)|x - z| \right] f_B^s dz \\ &+ \sum_{s \in E} N_W^s \times \int_{lb}^{ub} \left[(\hat{\alpha}_0^s + \hat{\alpha}_{BW}^s B) + (\hat{\beta}_0^s + \hat{\beta}_{BW}^s B)|x - z| \right] f_W^s dz\end{aligned} \tag{3}$$

where for each school s, N_B^s and N_W^s are the number of Black and White students, respectively. B and W are indicator variables taking the value 1 if a student is Black and White, respectively. $f_B^s(z)$ and $f_W^s(z)$ are the GPA distribution for Black and White students in school s. $\hat{\alpha}^s$ and $\hat{\beta}^s$ are the estimated parameters from equation (1) that capture the baseline probability of forming

[9] In the Appendix Figure A.8, we present simulation results using the coefficients from Table 2, and find comparable outcomes.

friendships and the interaction between GPA and race for the respective race pair. E represents the set of schools that contains all coefficients $\hat{\alpha}^s$ and $\hat{\beta}^s$. For low-achieving friends, lb and ub are designated as 1 and 3, respectively, while for high-achieving friends, lb and ub are set to 3 and 4.

The term $(\hat{\alpha}_0^s + \hat{\alpha}_{BB}^s B + \hat{\alpha}_{WB}^s W) + (\hat{\beta}_0^s + \hat{\beta}_{BB}^s B + \hat{\beta}_{WB}^s W)|x-z|$ represents the likelihood of forming friendships with Black peers when GPA distance is $|x-z|$. The term $(\hat{\alpha}_0^s + \hat{\alpha}_{BW}^s B) + (\hat{\beta}_0^s + \hat{\beta}_{BW}^s B)|x-z|$ represents the same likelihood for friendships with White peers. We calculate the $\tilde{n}$ for all GPAs x and compare them with the number of friends in the data.

We present the number of friends from the data in Figure 5 (a) and from simulation in Figure 5 (b). In Figure 5 (a), we present the number of friends by race (n_j^W and n_j^B) and by friends' achievement, where the achievement is divided into high and low types. As GPA increases, both Black and White students are more likely to receive friendship nominations from high-GPA students, but the slope is steeper for White students, creating a larger gap for students with high GPAs. Conversely, they are less likely to receive nominations from low-GPA students as GPA increases, but the change in the difference between Black and White students is not as significant as the change in nominations from high-GPA students.

Figure 5 (b) displays the number of simulated friends for Black and White students ($\tilde{n}^B$ and $\tilde{n}^W$) separately by friends' achievement. The distribution of the simulated number of friends captures both qualitatively and quantitatively the patterns of the distribution in the data. We also calculate the average number of friends by GPA ranges 1-2, 2-3, and 3-4 for both data and simulation in Appendix Table A.8.[10] Figure 5 (c) presents a comparison between data and simulation regarding the racial difference in the number of high- and low-achieving friends (i.e., $n_j^W - n_j^B$ and $\tilde{n}^W - \tilde{n}^B$). Regardless of students' GPA distribution and friends' achievement, the racial gap is similar for the data and simulation, confirming the model's goodness of fit. As detailed in Appendix A.6, the difference in the number of friends between our simulation and the data is consistently less than 0.2 across all GPA ranges.

[10] An observed dissimilarity between the data and simulation is that high-achieving students tend to have fewer friends in the simulation for both Black and White students. This discrepancy might indicate an unaccounted premium for high achievers, which our simple model does not consider. However, given our focus on racial comparisons, the omission of this factor, which affects both racial groups similarly, is unlikely to pose a significant issue in our racial comparison.

In the following subsections, using the simulated number of friends, we identify the factors contributing to the lower number of friends among high-achieving Black students compared to high-achieving White students by decomposing the gap into two factors: the population size and the GPA distribution.

4.1 Decomposing the Friendship Gap: Level and Composition Effects

We investigate the influence of changes in racial composition and GPA distribution on the racial friendship gaps, focusing on high-achieving students. To answer this, we compare the simulated gap before and after the distributional shifts among Black and White students relative to the gap at the lowest GPA. This analysis takes two steps. First, we equalize the share of Black and White students in each school by proportionally increasing the number of students to achieve equal proportions. This change is similar to the transition from Figure 2 (a) to (b). Second, in addition to equalization, we substitute the GPA distribution of White students for that of Black students, which is similar to aligning the GPA distribution of Black students with that of White students in 2 (b).

We calculate the racial friendship gap relative to the lowest GPA as follows:[11]

$$\begin{aligned} d(x,\hat{\alpha},\hat{\beta};N_B,N_W,f_B,f_W) =& \tilde{n}^W(x,\hat{\alpha},\hat{\beta};N_B,N_W,f_B,f_W) - \tilde{n}^B(x,\hat{\alpha},\hat{\beta};N_B,N_W,f_B,f_W) \\ & - \tilde{n}^W(x=1,\hat{\alpha},\hat{\beta};N_B,N_W,f_B,f_W) - \tilde{n}^B(x=1,\hat{\alpha},\hat{\beta};N_B,N_W,f_B,f_W). \end{aligned} \tag{4}$$

In Figure 6, we compare the simulated number of high- and low-achieving friends before and after distributional changes.[12] The solid blue line represents the difference in the simulated number of friends between White and Black students across the GPA distribution (i.e., $d(\hat{\alpha},\hat{\beta};N_B,N_W,f_B,f_W)$). Importantly, the friendship gap is mainly driven by high-achieving peers. High-achieving Black students have fewer high-achieving friends compared to their White counterparts, while they have more low-achieving friends. The red dashed line depicts the difference in the expected number of friends after Black and White student shares are equal-

[11] Without the normalization, the simulated number of friends for Black students at the lowest GPA in the counterfactual scenario of equal proportion with the same GPA distribution is larger than that for White students by 0.63 for high-achieving friends and 0.48 for low-achieving friends. See Appendix A.6 for further discussion and figures without the normalization.

[12] We present the overall gap including both high- and low-achieving friends in Appendix A.6.

ized (i.e., $d(\hat{\alpha},\hat{\beta};N^*,N^*,f_B,f_W)$, where $N^* = (N_B + N_W)/2$), and the black short dashed line represents the gap after substituting the White GPA distribution for the Black GPA distribution, as well as the equal shares (i.e., $d(\hat{\alpha},\hat{\beta};N^*,N^*,f_W,f_W)$).

Equalizing the shares of Black and White students in Figure 6 (a) narrows the gap in the number of friendship nominations from high achievers. Specifically, the gap decreases from 1.05 to 0.16 for students with a GPA of 3-4.[13] The reduced amount, 0.89, represents the level effect. We then substitute the White GPA distribution for the Black GPA distribution. After doing so, the gap decreases from 0.16 to 0.07. The reduced amount, 0.09, reflects the composition effect.[14] In total, the gap is reduced from 1.05 to 0.07 (a reduction of 93%).

In Figure 6 (b), we compare the difference in the expected number of friendships received from low achievers before and after distributional changes. After equalizing the shares, the gap increases from -0.26 to -0.01 for Black students with a GPA of 3-4. The increased amount, 0.25, reflects the level effect. We then compare the difference in the expected number of friendships received from low achievers before and after substituting the White GPA distribution for the Black GPA distribution. For a Black student with a GPA of 3-4, the gap decreases from -0.01 to -0.15. The reduced amount, 0.14, reflects the composition effect. In total, the gap is reduced from -0.26 to -0.15 (a reduction of 42%).

This analysis provides insights into the potential impact of altering the compositions of Black and White students on the expected number of friends, particularly for high-achieving Black students compared to their White counterparts. We find that high-achieving Black students have fewer friends than their White counterparts because of the smaller pool of high-achieving friends. Through our counterfactual analysis, we further find that most of the gap can be attributed to level effects. In other words, the smaller number of Black students' friends, particularly for high achievers, is not a result of the students' personal characteristics or culture but the circumstance of being a minority in school.

[13] In Appendix Table A.9, we present the numerical values of expected changes in the number of friends before and after composition changes by GPA.

[14] The dominance of level effects is due to the fact that increasing the proportion of Black students can significantly reduce the racial gap in student composition within the school. In Figure 2, much of the gap in the number of students, especially among high achievers, between Black and White students disappears when their proportions are equalized. Conversely, equalizing distribution without increasing the proportion does not result in significant changes in the racial composition of the school.

4.2 From the Friendship Gap to the Earnings Gap

In the preceding section, we established that the smaller number of friends of Black students primarily stems from their status as minorities within schools. In this section, we explore the labor market consequences of this reduced social connectivity.

We conduct a back-of-the-envelope calculation using our estimated reduction in the friendship gap in counterfactual scenarios and causal estimates of the earnings return to the number of friends. According to our counterfactual analysis, the expected friendship gap among high-achieving Black students decreases from 1.05 to 0.07 for high-achieving friends and from -0.26 to -0.15 for low-achieving friends. Thus, the total expected decreased gap in the number of friends for high-achieving students is 0.87 (i.e., 0.87 more friends for Black students).

Following Lleras-Muney et al. (2020), we conduct an instrumental variable (IV) analysis to estimate the causal effects of the number of friends, using the average absolute difference in age between a student and their peers in the same school and grade as the IV. In doing so, we exploit the fact that age difference is correlated with the probability of becoming friends (McPherson, Smith-Lovin, and Cook 2001) and uncorrelated with unobservable variations in earnings. For the earnings analysis, we include students of all races to maximize the number of observations, and restrict friends in the same grade to use the IV. Connecting students to their earnings and other personal traits, we regress the log yearly earnings of individual i in school j in grade g on the number of friends:

$$ln(earnings)_i = \gamma_0 + \gamma_b Black_i + \gamma_f NumFriend_i + \iota X_{ijg} + \delta_j + \eta_g + e_i \tag{5}$$

where *Black* is an indicator denoting Black students, and *NumFriend* is the number of friends in the same grade. We include school fixed effects, δ_j, and grade fixed effects, η_g. We also include individual-level and school-by-grade-level controls, X_{ijg}, including age, sex, race, education, social skills, cognitive skills, and the school-by-grade averages of those variables.[15]

Table 3 presents OLS and 2SLS results, with first-stage results provided in Appendix Table A.11. While it is not presented in the table, we include an interaction term between the number

[15] See Appendix A.7 for details on variable description.

of friends (or high- and low-achieving friends) and a dummy variable for races other than Black and White.[16] Thus, the variable *NumFriend* represents estimation results for White individuals. In columns (1) and (2), we present OLS estimates for the return to the number of friends. Since our focus is the racial gap, we estimate returns separately for Black and White students. Column (2) further distinguishes between high- and low-achieving friends. We find no evidence of different returns to number of friends between Black and White students, regardless of whether friends are high achieving or low achieving. Additionally, we find suggestive evidence that the return for high-achieving friends is higher than that for low-achieving friends.[17]

In columns (3) and (4), IV regression estimates are presented using overall age distance and race-specific age distance as instruments, respectively. Overall age distance is calculated based on age distance among all possible pairs in a school-grade following Lleras-Muney et al. (2020). The race-specific distance is calculated separately for Black and White peers to account for racial homophily in friendship formation. In column (3), we estimate the overall return to one more friend for students of all races. In column (4), we estimate the return separately for Black and White students using race-specific distance as the IV. We find a larger return to the number of friends in both cases while results are not statistically significant at the 5% level because of large standard errors. Once again, we do not find a different return between Black and White students.[18] In their various specifications, Lleras-Muney et al. (2020) estimate the bounds of return to the number of friends as 6.50% to 13.67%. Since our estimates mostly fall within these bounds and are not precise, we use their bounds for our back-of-the-envelope calculation for the racial earnings gap.[19]

With the same basic controls as in equation (5), high-achieving Black students earn 28.7% less than their White counterparts, with average earnings of $37,300, and respondents are primarily aged between nineteen and twenty-five years old.[20] In our counterfactual scenario,

[16] Students of races other than Black and White show no difference in returns compared to White students. See Appendix A.12 for results for students of races other than Black and White.

[17] The instruments lack sufficient power to produce precise estimates for different types of friendships.

[18] In Appendix Table A.12, we also present results dividing students based on achievement. We find that high-achieving students gain a higher return from having one more friend, while the difference is not large.

[19] Lleras-Muney et al. (2020) use the same data source and define a friend in the same manner, but sample selection is different. Their point estimate under a similar specification to column (3) is 12.42%. See their Table 3 for more details.

[20] After controlling for basic demographics, including age, sex, race, social skills, cognitive skills, and school and grade fixed effects, the racial earnings gap falls to 20.3%.

we anticipate a 0.87-point decrease in the gap in the number of friends for high-achieving Black students, resulting in a 5.7% to 11.9% increase in their earnings. Consequently, the wage gap diminishes from 28.7% to 19.3%-24.2%, representing a reduction of the gap by 4.5-9.4 percentage points (15.7%-32.8%). The reduction would be lower-bound if the causal estimate of the return to high-achieving friends exceeds that of low-achieving friends, as suggested by the OLS estimates in column (2).

5 Conclusion

This paper sheds light on a little-discussed disadvantage faced by students in minority groups in school: Black students tend to have fewer friends, particularly high-achieving ones, because the pool of similar peers is limited. This finding challenges previous views attributing the friendship gap to ostracization of high-achieving minorities. Based upon the estimation results from a dyadic regression examining the interaction between race and achievement, we conducted a counterfactual analysis with an equal proportion of races and the same GPA distribution for Black and White students. Under this scenario, the majority of the gap in friendships for high achievers disappears.

Furthermore, we established a link from the gap in friendships to the earnings gap. Our findings indicate that the disparity in the number of friends can explain as much as a 5-9 percentage point difference in yearly earnings between Black and White individuals. Minority groups face significant disadvantages before entering the labor market even in the absence of discrimination.

One significant implication of this study, albeit not discussed explicitly, concerns school racial desegregation programs (Akbar et al. 2022; Ananat 2011; Angrist and Lang 2004; Cutler and Glaeser 1997; Mele 2020). Our work suggests that because of racial and achievement homophily, transporting minorities to White-majority schools might not necessarily reduce segregation within schools. Particularly high-achieving Black students may lose other high-achieving Black peers as a result of transportation, which is not easily compensated by having more high-achieving White students because of racial homophily.

While we underscore the adversities faced by minority groups in schools and emphasize that

they may extend into the labor market, we leave it for future research to address the challenges faced by minority students.

References

Akbar, Prottoy A, Sijie Li Hickly, Allison Shertzer, and Randall P Walsh. 2022. "Racial Segregation in Housing Markets and the Erosion of Black Wealth." *Review of Economics and Statistics,* 1–45.

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Chapter 48 Race and Gender in the Labor Market." In *Handbook of Labor Economics,* 3:3143–3259. Elsevier, January. https://doi.org/10.1016/S1573-4463(99)30039-0.

Ananat, Elizabeth Oltmans. 2011. "The Wrong Side (s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality." *American Economic Journal: Applied Economics* 3 (2): 34–66.

Andrews, Rodney J, and Omari H Swinton. 2014. "The Persistent Myths of "Acting White" and Race Neutral Alternatives to Affirmative Action in Admissions." *The Review of Black Political Economy* 41 (3): 357–371.

Anelli, Massimo, and Giovanni Peri. 2019. "The Effects of High School Peers' Gender on College Major, College Performance and Income." *The Economic Journal* 129 (618): 553–602.

Angrist, Joshua D, and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review* 94 (5): 1613–1634.

Arrow, Kenneth J., and Ron Borzekowski. 2004. *Limited Network Connections and the Distribution of Wages.* SSRN Scholarly Paper. Rochester, NY, August. https://doi.org/10.2139/ssrn.632321.

Austen-Smith, David, and Roland G Fryer Jr. 2005. "An Economic Analysis of "acting White"." *The Quarterly Journal of Economics* 120 (2): 551–583.

Badev, Anton. 2021. "Nash Equilibria on (un) Stable Networks." *Econometrica* 89 (3): 1179–1206.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940." *The Quarterly Journal of Economics* 133, no. 3 (August): 1459–1501. https://doi.org/10.1093/qje/qjy003.

Bifulco, Robert, Jason M Fletcher, and Stephen L Ross. 2011. "The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health." *American Economic Journal: Economic Policy* 3 (1): 25–53.

Brown, Charles. 1984. "Black-white Earnings Ratios Since the Civil Rights Act of 1964: The Importance of Labor Market Dropouts*." *The Quarterly Journal of Economics* 99, no. 1 (February): 31–44. https://doi.org/10.2307/1885719.

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen. 2019. "Cool to Be Smart or Smart to Be Cool? Understanding Peer Pressure in Education." *The Review of Economic Studies* 86 (4): 1487–1526.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *The quarterly journal of economics* 130 (3): 1329–1367.

Calvó-Armengol, Antoni, and Matthew O. Jackson. 2004. "The Effects of Social Networks on Employment and Inequality." *American Economic Review* 94, no. 3 (June): 426–454. https://doi.org/10.1257/0002828041464542.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. "Peer Effects and Social Networks in Education." *The review of economic studies* 76 (4): 1239–1267.

Carrell, Scott E, Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review* 108 (11): 3377–3415.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, et al. 2022a. "Social Capital I: Measurement and Associations with Economic Mobility." *Nature* 608 (7921): 108–121.

———. 2022b. "Social Capital Ii: Determinants of Economic Connectedness." *Nature* 608 (7921): 122–134.

Coleman, James S. 1961. "The Adolescent Society."

Currarini, Sergio, Matthew O Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica* 77 (4): 1003–1045.

———. 2010. "Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation." *Proceedings of the National Academy of Sciences* 107 (11): 4857–4861.

Cutler, David M, and Edward L Glaeser. 1997. "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics* 112 (3): 827–872.

Diamond, John B, and James P Huguley. 2014. "Testing the Oppositional Culture Explanation in Desegregated Schools: The Impact of Racial Differences in Academic Orientations on School Performance." *Social Forces* 93 (2): 747–777.

Eguia, Jon X. 2017. "Discrimination and Assimilation at School." *Journal of Public Economics* 156:48–58.

Flashman, Jennifer. 2012a. "Academic Achievement and Its Impact on Friend Dynamics." *Sociology of education* 85 (1): 61–80.

———. 2012b. "Different Preferences or Different Opportunities? Explaining Race Differentials in the Academic Achievement of Friends." *Social Science Research* 41 (4): 888–903.

Fryer Jr, Roland G, and Paul Torelli. 2010. "An Empirical Analysis of 'acting White'." *Journal of Public Economics* 94 (5-6): 380–396.

Hanselman, Paul, Sarah K Bruch, Adam Gamoran, and Geoffrey D Borman. 2014. "Threat in Context: School Moderation of the Impact of Social Identity Threat on Racial/ethnic Achievement Gaps." *Sociology of Education* 87 (2): 106–124.

Holzer, Harry J. 1987. "Informal Job Search and Black Youth Unemployment." *The American Economic Review* 77 (3): 446–452. https://www.jstor.org/stable/1804107.

Imberman, Scott A, Adriana D Kugler, and Bruce I Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102 (5): 2048–2082.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101, no. 3 (June): 410–442. https://doi.org/10.1086/261881.

Lang, Kevin, and Jee-Yeon K. Lehmann. 2012. "Racial Discrimination in the Labor Market: Theory and Empirics." *Journal of Economic Literature* 50, no. 4 (December): 959–1006. https://doi.org/10.1257/jel.50.4.959.

Lleras-Muney, Adriana, Matthew Miller, Shuyang Sheng, and Veronica T Sovero. 2020. *Party on: The Labor Market Returns to Social Networks and Socializing*. Technical report. National Bureau of Economic Research.

Lundberg, Shelly. 2013. "The College Type: Personality and Educational Inequality." *Journal of Labor Economics* 31 (3): 421–441.

Marmaros, David, and Bruce Sacerdote. 2006. "How Do Friendships Form?" *The Quarterly Journal of Economics* 121 (1): 79–119.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American sociological review,* 122–131.

Mayer, Adalbert, and Steven L Puller. 2008. "The Old Boy (and Girl) Network: Social Network Formation on University Campuses." *Journal of public economics* 92 (1-2): 329–347.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual review of sociology* 27 (1): 415–444.

Mele, Angelo. 2020. "Does School Desegregation Promote Diverse Interactions? an Equilibrium Model of Segregation Within Schools." *American Economic Journal: Economic Policy* 12 (2): 228–257.

Nakajima, Ryo. 2007. "Measuring Peer Effects on Youth Smoking Behaviour." *The Review of Economic Studies* 74 (3): 897–935.

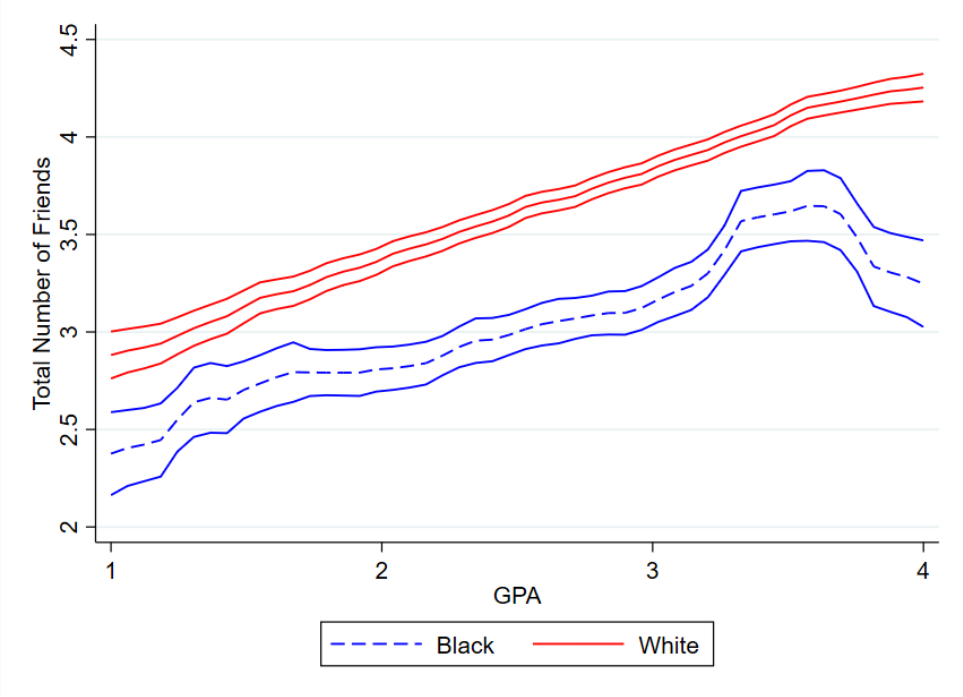
Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104, no. 5 (October): 869–895. https://doi.org/10.1086/262045.

Smirnov, Ivan, and Stefan Thurner. 2017. "Formation of Homophily in Academic Performance: Students Change Their Friends Rather Than Performance." *PloS one* 12 (8): e0183473.

Tyson, Karolyn, and Amanda E Lewis. 2021. "The "burden" of Oppositional Culture Among Black Youth in America." *Annual Review of Sociology* 47:459–477.

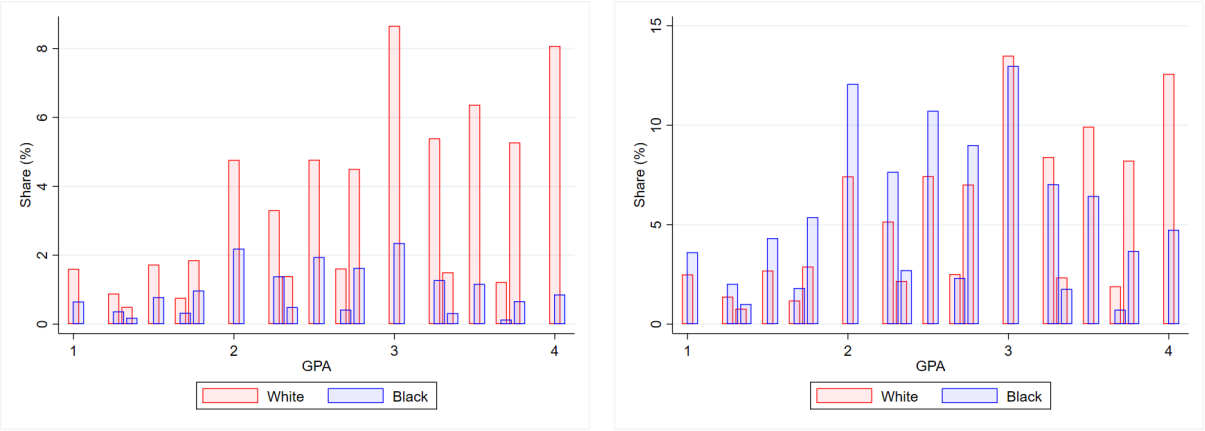
6 Figures and Tables

Fig. 1. Number of Friends by GPA



Notes. The figure depicts the locally smoothed means of the number of friends across GPA distribution separately by Black and White students.

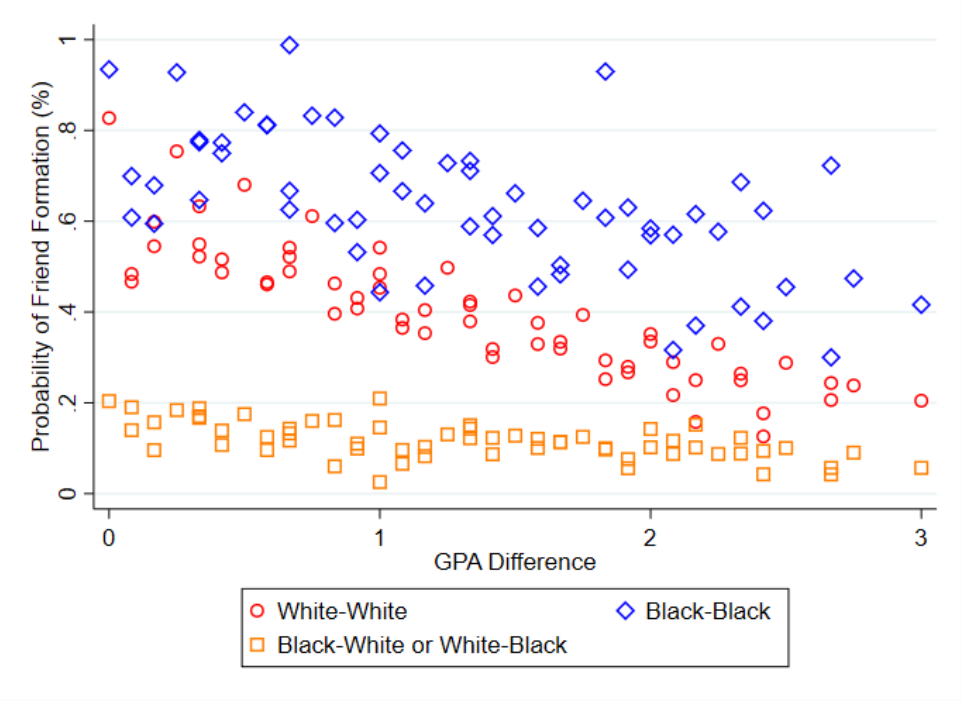
Fig. 2. GPA Distribution by Race



(a) Original Shares **(b)** Equal Shares

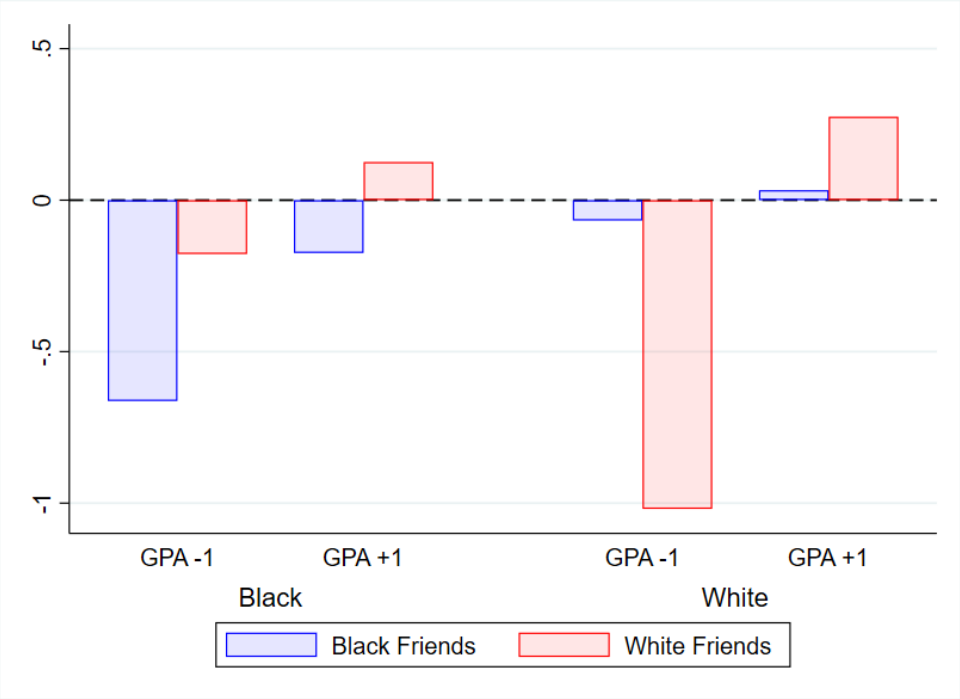
Notes. The figures illustrate the distributions of average GPA for Black and White students. The red bar indicates the density of the average GPA of White students. The blue bar indicates the density of the average GPA of Black students. Figure (a) illustrates the average GPA distribution with original Black and White shares. Figure (b) illustrates the average GPA distribution after equalizing Black and White shares.

Fig. 3. Probability of Forming Friendships by GPA Difference and Race



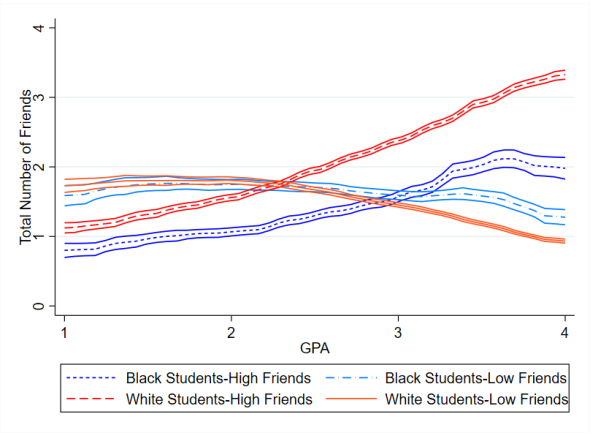
Notes. The figure illustrates the probability of forming friendships based on the GPA difference between pairs, depicted separately for racial combinations. Each potential dyad of students in each school is treated as an observation.

Fig. 4. Difference in the Number of Expected Friends Between Students with Average GPA and Students with Higher/Lower GPA

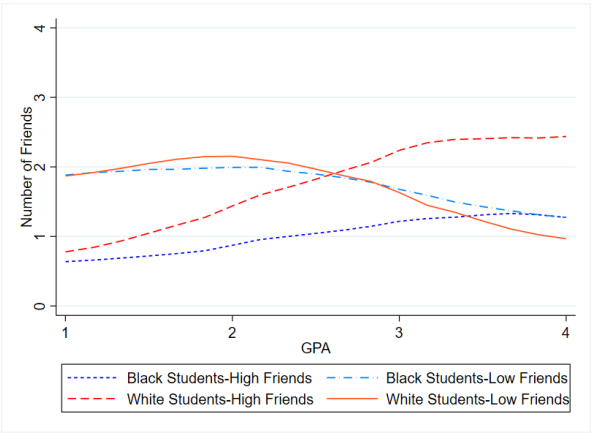


Notes. The figure illustrates the difference in the number of simulated friendships between students with the average GPA of Black students and students with a GPA that is higher or lower by 1 than this average. The blue and red bars represent the differences in Black and White friends, respectively. The expected numbers are computed using equation (1).

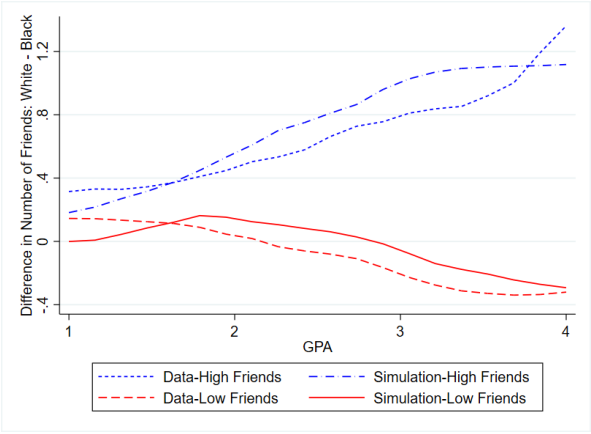
Fig. 5. Number of Friends from Data and Simulation



(a) Data



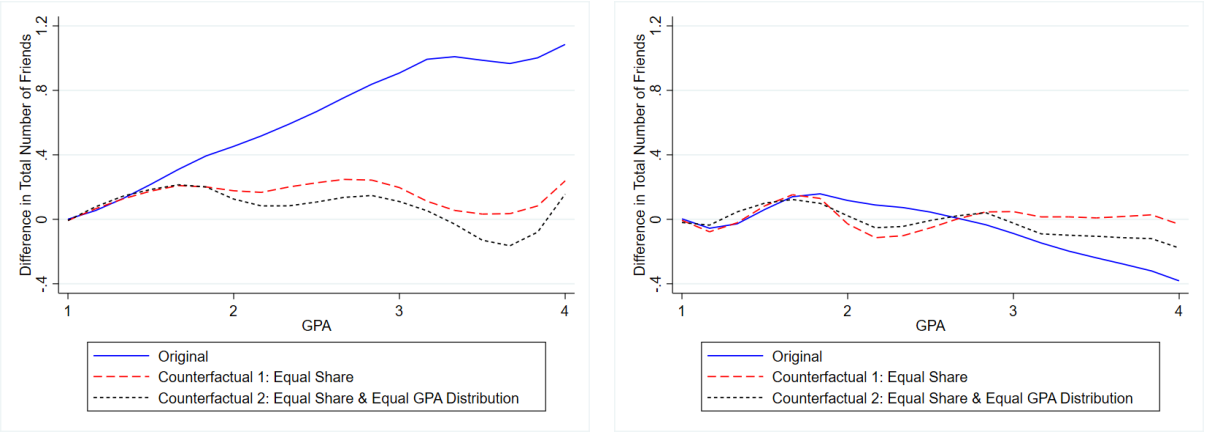
(b) Simulation



(c) Data-Simulation Comparison: Racial Friendship Gap

Notes. Figure (a) shows the number of friends across the GPA distribution, separately by race (Black and White) of the receiver and achievement (high and low) of the sender. High-achieving students are defined as those with a GPA of at least 3. The short dashed blue line and dashed-dotted blue line represent the number of high- and low-achieving friends of Black students, respectively. The dashed red line and solid red line represent the number of high- and low-achieving friends of White students, respectively. Figure (b) shows the simulated number of friends for Black and White students, differentiated by the sender's achievement following equation (3). The line styles in Figure (b) correspond to those in Figure (a). Figure (c) compares the Black and White friendship gap observed in the data (Figure (a)) with the corresponding simulation results (Figure (b)). This comparison is done separately for high- and low-achieving friends. The short dashed blue line and dashed-dotted blue line indicate the racial gap in the number of high-achieving friends from the data and simulation, respectively. The dashed red line and the solid red line indicate the racial gap in the number of low-achieving friends from the data and simulation, respectively.

Fig. 6. Racial Difference (White - Black) in the Number of Simulated Friends by GPA by High- and Low-Achieving Friends



(a) High-achieving Friends **(b)** Low-achieving Friends

Notes. The figure illustrates the difference in the number of expected friends between Black and White students, standardized to 0 at the lowest GPA level following equation (4). The simulated numbers are computed using equation (3). Figures (a) and (b) present the difference in the number of high-achieving and low-achieving friends, separately. The solid blue line depicts the original difference in the simulated number of friends. The dashed red line represents the difference assuming an equal proportion of Black and White students. The short-dashed black line represents the difference assuming both an equal share of Black and White students and an identical GPA distribution for Black and White students.

Table 1: Summary, Pairwise Level

	White→White	Black→Black	White→Black	Black→White	All
Friend (%)	0.567	0.784	0.125	0.103	0.513
GPA diff.	0.9	0.8	0.9	0.9	0.9
Share (%)	73.29	9.91	8.40	8.40	100
Observations	26,517,518	3,584,014	3,040,862	3,040,862	36,183,256

Notes. Friend(%) measures the shares of friendship formation out of all potential pairs of students. If student i nominates student j as a friend but j does not reciprocate, we define the $i-j$ pair but not the pair $j-i$ as friends. GPA difference measures the average difference in GPA between pairs.

Table 2: Friendship Formation by Race and Achievement

	Dependent variable : *i* nominating *j* as a friend (%)				
	(1)	(2)	(3)	(4)	(5)
dGPA	-0.189***	-0.164***	-0.216***	-0.189***	-0.259***
	(0.019)	(0.015)	(0.025)	(0.022)	(0.028)
$dGPA \times BB$			0.047	0.017	0.038
			(0.044)	(0.045)	(0.049)
$dGPA \times BW$			0.188***	0.158***	0.200***
			(0.025)	(0.022)	(0.025)
$dGPA \times WB$			0.182***	0.158***	0.193***
			(0.024)	(0.023)	(0.026)
BB			0.168	0.250	
			(0.166)	(0.171)	
BW			-0.624***	-0.627***	
			(0.078)	(0.061)	
WB			-0.597***	-0.603***	
			(0.078)	(0.060)	
Constant	0.674***	0.694***	0.752***	0.862***	0.938***
	(0.060)	(0.097)	(0.080)	(0.129)	(0.051)
Observations	36,183,256	36,183,256	36,183,256	36,183,256	36,183,256
Mean	0.513	0.513	0.513	0.513	0.513
Controls		X		X	X
Individual FE					X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student *i* nominates student *j* as a friend. *dGPA* indicates a difference in average GPA within a pair of students. *B* indicates a Black student, and *W* indicates a White student. *BW* is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects. Individual FE indicates that individual fixed effects of senders and receivers are controlled for. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

Table 3: Effects of the Number of Friends on Earnings

	OLS		2SLS	
	(1)	(2)	(3)	(4)
Friends	0.024***		0.146*	0.088
	(0.004)		(0.081)	(0.092)
Friends× Black	-0.002			-0.006
	(0.012)			(0.069)
High-Achieving Friends		0.029***		
		(0.004)		
Low-Achieving Friends		0.011*		
		(0.006)		
High-Achieving Friends× Black		0.003		
		(0.013)		
Low-Achieving Friends× Black		-0.005		
		(0.014)		
Controls	X	X	X	X
Sample	All	All	All	All
IV			Overall	Racial
Observations	9059	9059	9059	9059

Notes. The table presents estimation results from equation (5). Columns (1) and (2) present OLS estimates and columns (3) and (4) present 2SLS estimates in which the number of friends is endogenous. See A.7 for first-stage results. “Controls" stands for the inclusion of control variables, including age, sex, race, social skills, cognitive skills, and school and grade fixed effects. “Overall” indicates average age distance for all students in a school-grade. “Racial" indicates the race-specific average age distance in school-grade. We use students of all races. To facilitate the interpretation of the “Friends" variables, we include an interaction term between the number of friends (or high- and low-achieving friends) and a dummy variable for races other than Black and White. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

ONLINE APPENDIX

The Racial Gap in Friendships Among High-Achieving Students

Weonhyeok Chung and Jeonghyeok Kim

Part

Appendix

Table of Contents

A Additional Results

A.1 Summary Statistics

Table A.1: Summary, Individual Level

	White	Black	All
GPA	2.9	2.6	2.8
Female (%)	51	55	52
Age	15	15	15
Number of Friends	3.80	3.04	3.64
High Performing Friends (GPA$\geq$3)	2.38	1.40	2.17
Low Performing Friends (GPA$<$3)	1.42	1.64	1.47
Share (%)	78.9	21.1	100
Observations	43,109	11,530	54,639

A.2 Calculating Probability of Forming Friends

The following illustration outlines the methodology employed to generate Figure 3. Consider a school with a total of 15 students. Within this group, there are two high-achieving Black students, three low-achieving Black students, five high-achieving White students, and five low-achieving White students.

Table A.2: Example with Pairs of 15 Students

i and j		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		B_H	B_H	B_L	B_L	B_L	W_H	W_H	W_H	W_H	W_H	W_L	W_L	W_L	W_L	W_L
1	B_H	-	1	1	1	0	0	0	0	0	1	0	0	0	0	1
2	B_H	1	-	0	1	1	0	1	0	0	0	0	0	0	0	0
3	B_L	1	0	-	1	1	0	0	0	0	1	0	0	0	0	0
4	B_L	0	0	1	-	0	0	0	0	0	0	0	0	0	1	0
5	B_L	1	1	1	1	-	0	0	0	0	1	1	0	1	0	0
6	W_H	0	0	0	0	0	-	1	1	1	0	1	0	0	0	1
7	W_H	0	0	0	0	0	1	-	1	0	1	0	0	0	0	0
8	W_H	0	0	0	0	0	1	1	-	0	0	1	0	0	0	1
9	W_H	0	0	0	0	0	1	0	0	-	1	0	0	0	0	0
10	W_H	1	1	1	0	1	0	1	1	0	-	1	0	0	0	1
11	W_L	0	0	0	0	1	1	0	1	1	1	-	1	0	1	0
12	W_L	0	0	0	0	0	0	0	0	0	0	0	-	1	0	0
13	W_L	0	0	0	0	0	0	0	0	0	0	1	0	-	1	0
14	W_L	0	0	0	1	0	0	0	0	0	0	0	0	1	-	1
15	W_L	1	0	0	0	0	1	0	1	0	1	1	1	0	1	-

Notes. This is a simple example with Black and White students with either high ability or low ability. Each row indicates a student who lists another student and each column indicates the student who are listed.

In Appendix Table A.2, we present the friendship connections among the 15 students. Each cell indicates whether student i in the row directs student j in the column as a friend. The labels in the table represent the race and achievement of the students, denoted as R_a, where R represents the race and a represents the achievement of the student. For instance, B_H in the first row (student $i = 1$) represents a high-achieving Black student. In the table, we can observe that student 2, who possesses a high-achievement, directs student 7, who possesses a high-achievement.

In Appendix Table A.3, we calculate the probabilities of forming friendships based on the data in Appendix Table A.2. For instance, consider the case of a high-achieving White student as the friendship sender and a low-achieving Black student as the friendship receiver. In this

Table A.3: Probability of Forming Friendships

	B_H	B_L	W_H	W_L
B_H	100%	66.7%	20%	10%
B_L	50%	83.3%	13.3%	20%
W_H	20%	13.3%	60%	24%
W_L	10%	13.3%	28%	50%

Notes. We calculate probabilities of forming friendship for each possible race and achievement based on Appendix Table A.2.

scenario, there are 15 potential pairs (5 high-achieving White students and 3 low-achieving Black students) that can form a friendship. However, only 2 actual friendships are formed out of these potential pairs. Therefore, the probability of a high-achieving White student directing a low-achieving Black student as a friend is $\frac{2}{15} = 13.3\%$.

Fig. A.1. Probability of Forming Friendships by GPA Difference and Race

Notes. The figure illustrates average probability of friendship formation by GPA difference and race pairs based on the Appendix Table A.2. In this figure, GPA difference is 0 when two students are both high type or low type. GPA difference is 1 when two students are different achievement type (high-low or low-high types).

Using Appendix Table A.2, we plot the relationship between the GPA difference and the probability of friendship formation for each possible pair in Appendix Figure A.1. For instance, there are two cases where two Black students have identical GPAs: a high-achieving Black student directing another high-achieving Black student as a friend, and a low-achieving Black student directing another low-achieving Black student as a friend. In the figure, the likelihood of a high-achieving Black student directing another high-achieving Black student as a friend is

100%, while the likelihood of a low-achieving Black student directing another low-achieving Black student as a friend is 83.3%.

Compared to these equally-achieving friend pairs, the likelihood of forming a friendship decreases when students have different GPAs. The likelihood of forming a friendship between Black students with different achievement is lower than that of the equally-achieving cases. Additionally, the likelihood of forming a friendship for same-race friendships is higher than that of different-race friendships for a given GPA difference.

A.3 Heterogeneity Across Black Shares in Schools

In this subsection, we explore heterogeneity in friendship formation across schools with different racial compositions. Racial composition may also modify the relationship between race and achievement as factors in friendship formation. We divide schools into four groups so each group has the same number of Black students: the first, second, third, and fourth groups include schools with Black student shares ranging from 0% to 22%, 23% to 40%, 41% to 74%, and 75% to 100%, respectively.

In Appendix Figure A.2, we first present the number of friends, categorized by race and the share of Black students. A few patterns are worth noting. High-achieving Black students in Black-minority schools (0%-40%) tend to have significantly fewer friends than their White peers, whereas this disparity is absent in schools in which the Black population is not a minority (41%-100%). Low-achieving Black students in Black-minority schools also exhibit differences in the number of friends compared to their White counterparts, although the difference is smaller than that observed among high-achieving students. Moreover, we present the GPA distribution of Black and White students by the proportion of Black students in Appendix Figure A.3. While White students tend to cluster above a GPA of 3 and Black students are concentrated between 2 and 3, Black-minority schools show more pronounced differences in the distribution of GPAs.[A.1] In essence, the figure suggests that high-achieving Black students tend to have a smaller circle of friends compared to their White peers, especially in schools in which the proportion of Black students is smaller, particularly at higher GPAs.

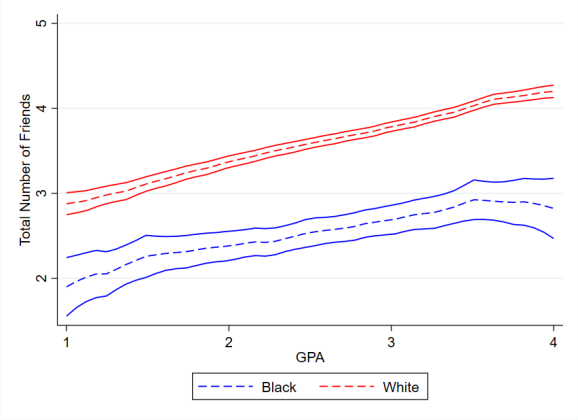
In Appendix Table A.4, we estimate equation (1) separately for students enrolled in schools with different shares of Black students. This analysis enables us to explore how friendship-formation patterns vary across shares of Black students. We present the estimation results starting with all schools in column (1), followed by schools categorized based on different shares of Black students in columns (2) to (5).

Across all types of schools, the significance of GPA differences is either similar or less pronounced for Black students compared to White students. Specifically, in columns (2) and

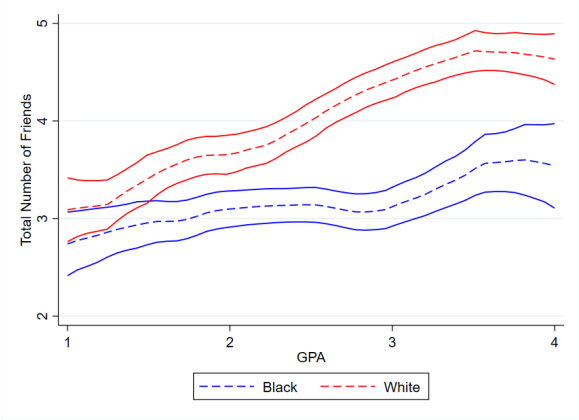
[A.1] Specifically, in schools with Black student shares ranging from 0%-22%, 23%-40%, 41%-74%, and 75%-100%, the gaps in shares with GPA of 3-4 between Black and White students are 19%, 19%, 13%, and 1%,respectively.

(3), we find no significant relationship between $dGPA \times BB$ (difference in GPA multiplied by an indicator for Black-student pair) and friendship formation in schools in which White students are in the majority. However, positive results are observed for schools with a Black student share exceeding 40%. Notably, the degree of overall racial homophily, denoted as BB, is contingent upon the proportion of Black students. Specifically, as the share of Black students increases, the corresponding coefficient diminishes. In Appendix Table A.5, we find similar results while controlling for friendship senders' and friendship receivers' fixed effects.

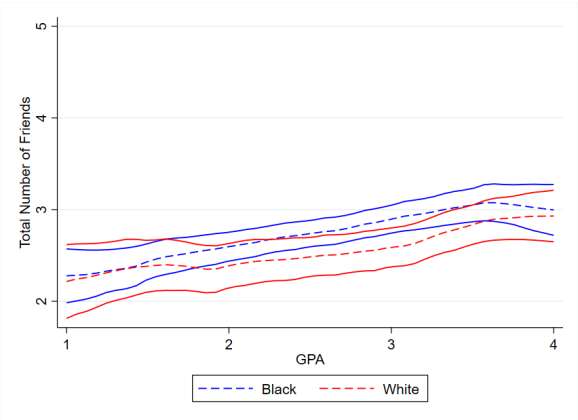
Fig. A.2. Number of Friends by GPA and Black Students' Share in Schools



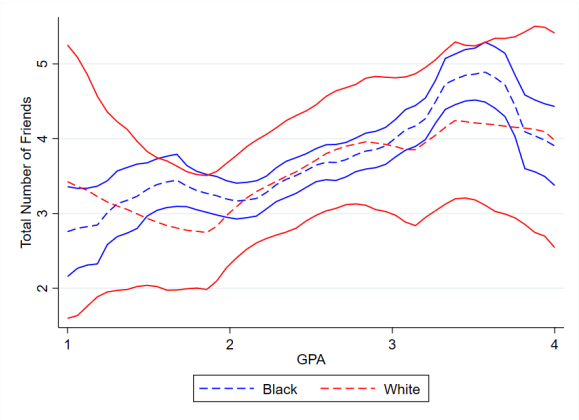
(a) Schools with 0%-22% Black enrollment



(b) Schools with 23%-40% Black enrollment



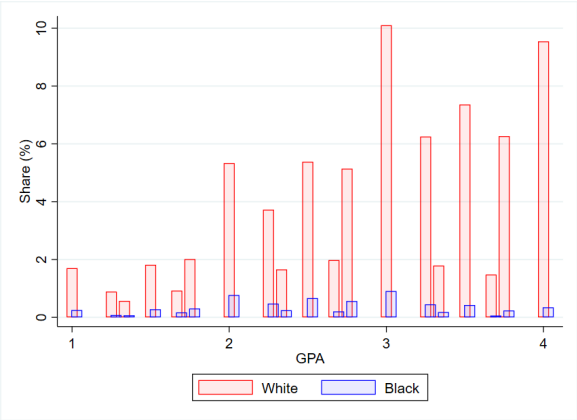
(c) Schools with 41%-74% Black enrollment



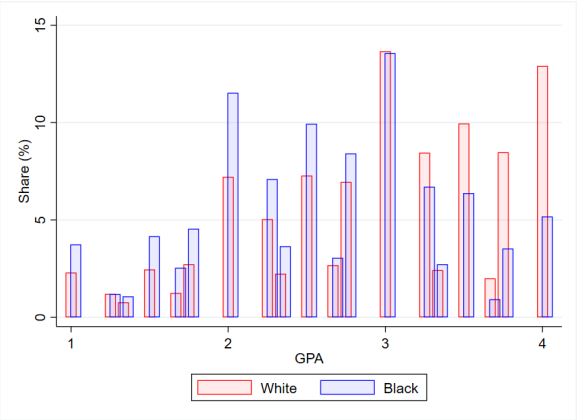
(d) Schools with 75%-100% Black enrollment

Notes. The figure depicts the locally smoothed means of the number of friends across GPA distribution, categorized by race and the share of Black students. We use Black and White students in the in-school survey of the National Longitudinal Study of Adolescent to Adult Health.

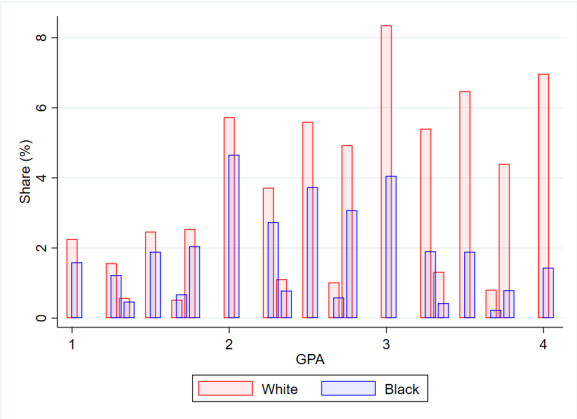
Fig. A.3. GPA Distribution by Race



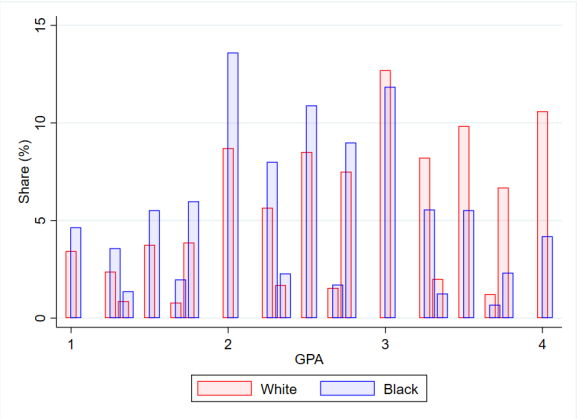
(a) Original Shares (Black 0%-22%)



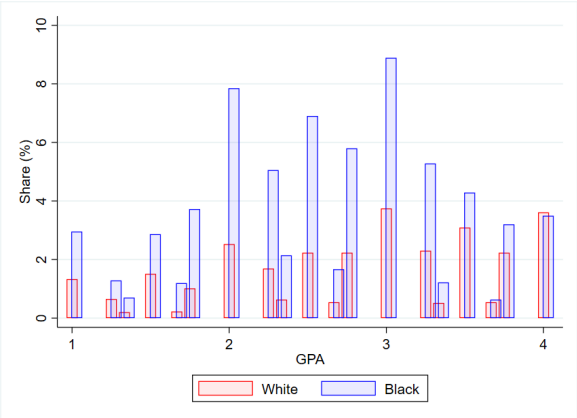
(b) Equal Shares (Black 0%-22%)



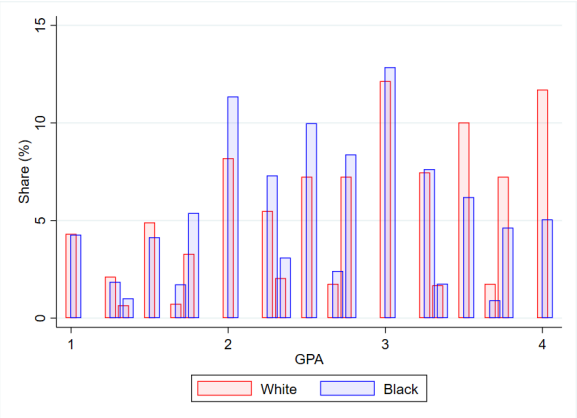
(c) Original Shares (Black 23%-40%)



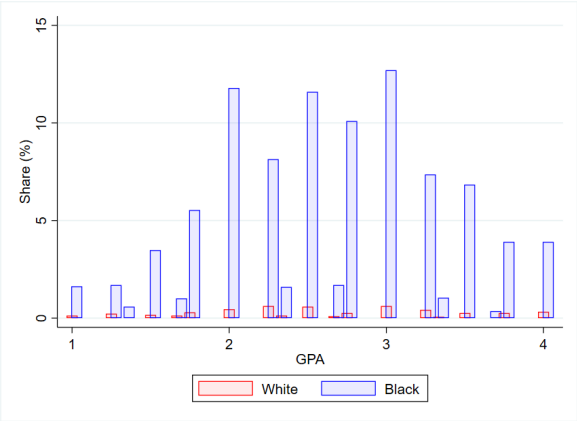
(d) Equal Shares (Black 23%-40%)



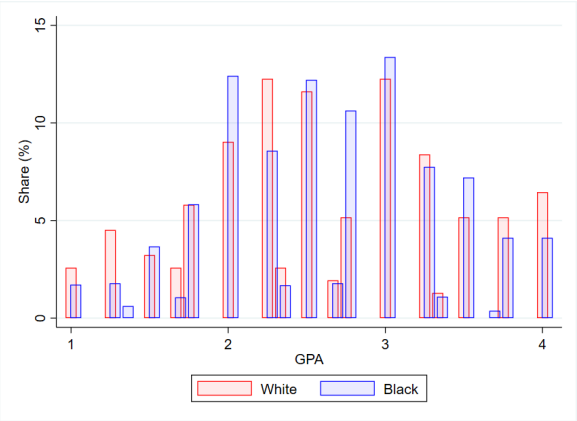
(e) Original Shares (Black 41%-74%)



(f) Equal Shares (Black 41%-74%)



(g) Original Shares (Black 75%-100%)



(h) Equal Shares (Black 75%-100%)

Notes. The figures illustrate the distributions of average GPA for Black and White students, separately by proportion of Black students in school. The red bar indicates the density of the average GPA of White students. The blue bar indicates the density of the average GPA of Black students. The sub-figures (a), (c), (e), and (g), illustrate the average GPA distribution with original Black and White shares. The sub-figures (b), (d), (f), and (h), illustrate the average GPA distribution after equalizing Black and White shares.

Table A.4: Friendship Formation by Race and Achievement: Across Black Student Share

	Dependent variable : Probability of i nominating j as a friend (%)				
	(1) All	(2) 0-22%	(3) 23-40%	(4) 41-74%	(5) 75-100%
$dGPA$	-0.189***	-0.179***	-0.284***	-0.267***	-1.361***
	(0.022)	(0.022)	(0.062)	(0.032)	(0.116)
$dGPA \times BB$	0.017	-0.095	0.044	0.187***	1.152***
	(0.045)	(0.075)	(0.032)	(0.032)	(0.127)
$dGPA \times BW$	0.158***	0.144***	0.267***	0.239***	1.317***
	(0.022)	(0.022)	(0.061)	(0.040)	(0.106)
$dGPA \times WB$	0.158***	0.145***	0.265***	0.245***	1.322***
	(0.023)	(0.022)	(0.063)	(0.034)	(0.098)
BB	0.250	1.212***	0.300**	-0.513***	-5.207***
	(0.171)	(0.285)	(0.128)	(0.122)	(0.186)
BW	-0.627***	-0.474***	-1.000***	-0.983***	-5.927***
	(0.061)	(0.048)	(0.208)	(0.144)	(0.223)
WB	-0.603***	-0.448***	-0.987***	-0.952***	-5.838***
	(0.060)	(0.047)	(0.213)	(0.128)	(0.259)
Constant	0.862***	0.741***	0.935***	1.701***	6.103***
	(0.129)	(0.098)	(0.276)	(0.244)	(0.301)
Observations	36,183,256	26,910,900	4,993,870	2,743,852	1,534,634
Mean	0.513	0.509	0.539	0.400	0.686
Controls	X	X	X	X	X

Notes. Notes. The table presents coefficients and standard errors from estimation of equation (1), separately by shares of Black students. Units of observation are directed dyads, based on non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominated student j as a friend. *dGPA* indicates a difference in average GPA within pairs of students. B indicates Black students, and W indicates White students. The schools are divided into four quantile groups as follows: the first quantile (Q1) includes schools with Black student shares ranging from 0% to 22%; the second quantile (Q2), 23% to 40%; the third quantile (Q3), 41% to 74%; and the fourth quantile (Q4), 75% to 100%. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

Table A.5: Friendship Formation by Race and Achievement: Across Black Student Proportion with Individual Fixed-Effects

	Dependent variable : Probability of i nominating j as a friend (%)				
	(1) All	(2) 0-22%	(3) 23-40%	(4) 41-74%	(5) 75-100%
$dGPA$	-0.259***	-0.248***	-0.359***	-0.269***	-1.406***
	(0.028)	(0.029)	(0.078)	(0.035)	(0.124)
$dGPA \times BB$	0.038	-0.080	0.094**	0.146***	1.136***
	(0.049)	(0.075)	(0.039)	(0.031)	(0.120)
$dGPA \times BW$	0.200***	0.165***	0.295***	0.221***	1.298***
	(0.025)	(0.025)	(0.067)	(0.039)	(0.106)
$dGPA \times WB$	0.193***	0.154***	0.281***	0.229***	1.354***
	(0.026)	(0.024)	(0.067)	(0.037)	(0.097)
Constant	0.938***	1.154***	0.870***	0.097	-4.600***
	(0.051)	(0.070)	(0.056)	(0.111)	(0.244)
Observations	36,183,256	26,910,900	4,993,870	2,743,852	1,534,634
Mean	0.513	0.509	0.539	0.400	0.681

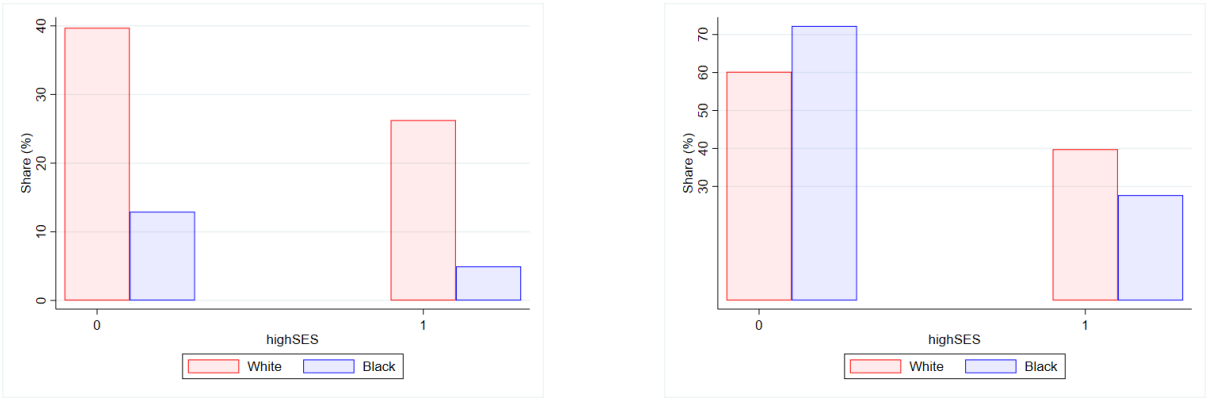
Notes. The table presents coefficients and standard errors from estimation of equation (1), separately by shares of Black students. Units of observation are directed dyads, based on non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes on the value of 1 if student i nominated student j as a friend. $dGPA$ indicates a difference in average GPA within the pairs of students. B indicates Black students, and W indicates White students. The schools are divided into four quantile groups as follows: the first quantile (Q1) includes schools with Black student shares ranging from 0% to 22%; the second quantile (Q2), 23% to 40%; the third quantile (Q3), 41% to 74%; and the fourth quantile (Q4), 75% to 100%. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

A.4 Mother's Years of Schooling or SES

In this subsection, we analyze the distribution of socioeconomic status (SES) groups and investigate friendship formation based on race and SES groups. We define socioeconomic status using two indicators of family background: the mother's educational attainment and whether the student lives with both parents. The high SES group includes students whose mothers have attained some college education or higher and who live with both parents. The remaining students are classified into the low SES group. Our definition of the high SES group aligns with the most advantaged group in Lundberg (2013).[A.2]

By utilizing SES groups based on family backgrounds, we also address concerns about the reverse causality between GPA and friendship formations. Instead of achievement affecting friendship formation, friendship formation may affect academic achievements. However, the measurement of the SES group is not affected by this issue as the SES group is a predetermined characteristic of students.

Fig. A.4. SES Distribution by Race



(a) Original Share **(b)** Black-White Equal Share

Notes. The figure illustrates the distribution of SES for Black and White students. The sub-figure (a) illustrates the distribution of SES with original Black and White shares. The sub-figure (b) illustrates the distribution of SES after equalizing Black and White shares. High SES includes students in high maternal education and residence with both parents.

Appendix Figure A.4 illustrates the distribution of socioeconomic status (SES) for Black and White students. In sub-figure (a), White students comprise the majority in both SES categories

[A.2] Although Lundberg (2013) defines the most advantaged group as students with both biological parents and high maternal education, we use both parents instead as we cannot identify whether both parents of students are biological parents from the in-school data that we use for the analysis.

while Black students represent the minority. In sub-figure (b), the distribution of SES for Black and White students is displayed after equalizing the shares of Black and White students. The figure presents that 30 percent of Black students are in the high SES category, whereas 40 percent of White students fall into the high SES category. Consequently, within their respective racial groups, Black students have a smaller pool of potential high SES friends compared to White students.

Table A.6: Friendship Formation by Race and SES

	Dependent variable : i nominating j as a friend (%)				
	(1)	(2)	(3)	(4)	(5)
DifferentSES	-0.078***	-0.089***	-0.083***	-0.090***	-0.110***
	(0.014)	(0.013)	(0.019)	(0.013)	(0.015)
DifferentSES $\times BB$			-0.022	-0.038**	-0.005
			(0.042)	(0.017)	(0.032)
DifferentSES $\times BW$			0.074***	0.072***	0.051**
			(0.020)	(0.020)	(0.025)
DifferentSES $\times WB$			0.063***	0.073***	0.061***
			(0.021)	(0.020)	(0.023)
BB			0.211	0.267*	
			(0.129)	(0.145)	
BW			-0.516***	-0.546***	
			(0.063)	(0.053)	
WB			-0.495***	-0.525***	
			(0.063)	(0.053)	
Constant	0.564***	0.705***	0.626***	0.839***	0.878***
	(0.047)	(0.103)	(0.064)	(0.129)	(0.043)
Observations	30,156,736	30,156,736	30,156,736	30,156,736	30,156,736
Mean	0.530	0.530	0.530	0.530	0.530
Controls		X		X	X
Individual FE					X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported mothers' education, residency with both parents, and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominates student j as a friend. $DifferentSES$ is a binary variable equal to one if both students are in a different SES group. B indicates Black student and W indicates White student. BW is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicates controlling for demographic variables including gender and age of senders and receivers, age difference within pairs, and school fixed effects. Individual FE indicates that individual fixed effects of senders and receivers are controlled for. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

Appendix Table A.6 presents estimation results in equation (1) with and without covariates, using the SES group instead of GPA. In column (1), if the SES groups differ, the probability

of friendship formation between two students decreases by 0.078 percentage points when the overall mean is 0.53 percent. In column (2), when we control for additional characteristics, including dyad-level characteristics (age and gender pairs) and demographic variables related to the senders' or receivers' genders and ages, the coefficient becomes more pronounced. In column (2), the results indicate that when the SES group differs, the probability of friendship formation between two students in the school decreases by 0.089 percentage points.

In column (3), we find that different SES matters similarly between pairs of White students and pairs of Black students. In column (4), We find similar results to those in column (3) after controlling dyad-level and individual-level characteristics. In column (5), we control sender fixed effects and receiver fixed effects. Regardless of whether we control for individual fixed effects, we find that the difference in SES matters similarly in friendship formation for Black and White students. Moreover, the significance of the SES difference is more pronounced for within-race friendships compared to across-race friendships, regardless of the control variables.

A.5 Forming Friendships within School Clubs

In Appendix Table A.7, we examine whether the difference in GPA has varying implications for friendship formation when students are members of the same social club or not. The variable *Club* indicates whether the two students are both members of at least one social club. Our findings show that students who share membership in a club are more likely to form friendships compared to those who do not share club membership. Furthermore, we observe that the difference in GPA matters more on friendship formation for student pairs who are part of the same club ($Club \times dGPA$). In other words, even when students belong to the same club, the likelihood of forming friendships diminishes if there are larger differences in their GPAs. These findings address concerns that the formation of friendships is solely influenced by exposure rather than the active choice of students. In addition, we do not find a significant difference between White-White pairs and Black-Black pairs, regardless of their involvement in the social club ($Club \times dGPA \times BB$ and $dGPA \times BB$). Moreover, the significance of the achievement difference is more pronounced for within-race friendships compared to across-race friendships, regardless of the control variables.

Table A.7: Friendship Formation by Race and Achievement: Within School Clubs

	Dependent variable : i nominating j as a friend (%)	
	(1)	(2)
$Club \times dGPA$	-0.355***	-0.353***
	(0.046)	(0.041)
$Club \times dGPA \times BB$	0.103	0.116
	(0.095)	(0.085)
$Club \times dGPA \times BW$	0.283***	0.288***
	(0.046)	(0.044)
$Club \times dGPA \times WB$	0.285***	0.290***
	(0.045)	(0.042)
$Club \times BB$	-0.123	-0.106
	(0.230)	(0.199)
$Club \times BW$	-0.962***	-0.936***
	(0.109)	(0.107)
$Club \times WB$	-0.923***	-0.896***
	(0.107)	(0.104)
$Club$	1.242***	1.096***
	(0.110)	(0.098)
$dGPA$	-0.130***	-0.127***
	(0.016)	(0.015)
$dGPA \times BB$	0.012	0.002
	(0.030)	(0.032)
$dGPA \times BW$	0.113***	0.100***
	(0.016)	(0.016)
$dGPA \times WB$	0.108***	0.100***
	(0.016)	(0.016)
BB	0.207	0.228
	(0.125)	(0.140)
BW	-0.434***	-0.489***
	(0.057)	(0.052)
WB	-0.413***	-0.470***
	(0.057)	(0.052)
Constant	0.526***	0.592***
	(0.058)	(0.125)
Observations	36,183,256	36,183,256
Mean	0.5	0.5
Controls		X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominates student j as a friend. $dGPA$ indicates a difference in average GPA within a pair of students. $Club$ is a binary variable equal to one if both students were in at least one extracurricular activity. B indicates Black students, and W indicates White students. Demographic variables are controlled for including gender and age of senders and receivers, difference in age within pairs, and school fixed effects.Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

A.6 Additional Simulation Outcomes

We present supplementary results for simulation analysis. Appendix Table A.8 displays the number of friends categorized by students' GPA (1-2, 2-3, 3-4), students' race (Black, White), and friends' GPA (high-achieving, low-achieving). Furthermore, we illustrate the disparity in the number of friends between Black and White students, as well as the difference between data and simulation regarding the racial gap in friendship counts.

Appendix Table A.9 provides numerical values for the difference in the number of friends in each counterfactual analysis, complementing Figure 6.

Appendix Figure A.5 displays the total number of friends gap, including both high- and low-achieving friends.

Appendix Figure A.6 illustrates the friendship gap in the original simulation and each counterfactual scenario without normalization, where the gap is set to zero at the lowest GPA. In counterfactual 1, we equalize the proportions of Black and White students, resulting in the near elimination of the racial gap in high-achieving friends, while the gap in low-achieving friends becomes around negative 1.0. This suggests that Black students are expected to have the same number of high-achieving friends and approximately one more low-achieving friend compared to White students. In counterfactual 2, we further equalize their GPA distributions, presenting that Black students have approximately 0.5 more high- and low-achieving friends.

However, it is important to note that the more friends of Black students compared to White in the counterfactuals may be influenced by the higher level of racial homophily due to the smaller proportion of Black students. In our regression model (equation (1)), the overall level of racial homophily is determined by the probability of forming friendships among all possible pairs. Since the number of friends is constrained to a certain level (i.e., congestion), the overall level of homophily may be overestimated when the number of Black students is limited.[A.3] Therefore, even if our estimation suggests that Black students have more friends, we focus on relative levels in the baseline analysis where overall homophily cannot exert influence. If there is an overall difference between races in preferences for forming friendships, and thus Black students have

[A.3] For example, having 10 Black friends among 20 Black peers results in a probability of 0.5. If there are 40 Black students, a student would need to have 20 Black friends to achieve the same probability. However, it is unrealistic to expect individuals to have such a large number of friends due to time and resource constraints.

a higher number of friends in our counterfactual scenarios, the reduced gap in the number of friends from our counterfactuals may be underestimated.

In Appendix Figure A.8, we present results of counterfactual analyses using representative coefficients derived from all schools and schools with Black enrollment percentages between 23-40% and 41-74%, as shown in Table 2 and A.4. In all instances, we observe a reduction in the racial gap in the number of friends when the Black-White proportion and GPA distributions are equalized. Notably, the explained gap for high-achieving students is larger than the baseline results in all cases. This mitigates concerns regarding the potential impact of student distributions on coefficients. Overall, we consistently attempt to adopt a more conservative approach when estimating the reduction in the friends gap in our counterfactual analyses.

Table A.8: Number of friends by Type of Friends and GPA

	Black Students		White Students		Difference (White - Black)	
	(1) Low-Perform Friends	(2) High-Perform Friends	(3) Low-Perform Friends	(4) High-Perform Friends	(5) Low-Perform Friends	(6) High-Perform Friends
Data						
GPA 1-2	1.71	0.98	1.79	1.37	0.08	0.39
GPA 2-3	1.66	1.37	1.60	2.10	-0.06	0.73
GPA 3-4	1.51	1.98	1.10	3.07	-0.41	1.09
Simulation Results						
GPA 1-2	1.98	0.77	2.08	1.18	0.10	0.41
GPA 2-3	1.87	1.11	1.87	2.00	0.00	0.89
GPA 3-4	1.44	1.28	1.15	2.44	-0.29	1.16
Difference (Data - Simulation)						
GPA 1-2	-0.27	0.21	-0.29	0.19	-0.02	-0.02
GPA 2-3	-0.21	0.26	-0.27	0.10	-0.06	-0.16
GPA 3-4	0.07	0.70	-0.05	0.63	-0.12	-0.07

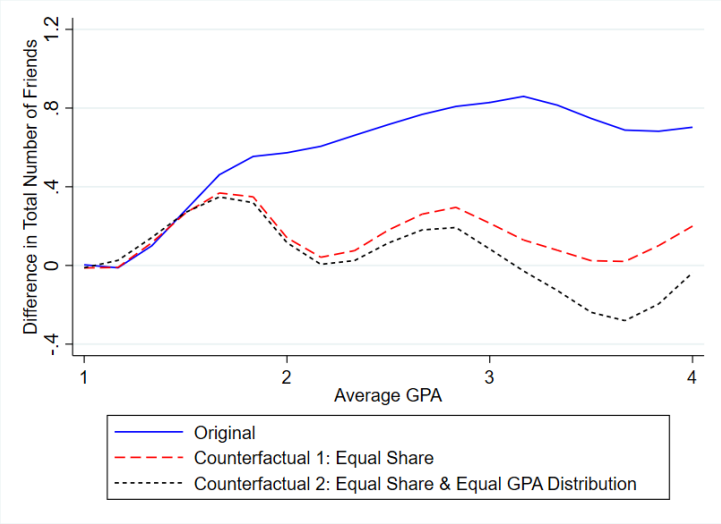
Notes. This table illustrates the number of friends based on the race of receivers and the achievement types of both receivers and senders. Receiver achievement is categorized into GPA ranges 1-2, 2-3, and 3-4, while sender achievement is divided into low (1-3) and high (3-4). The data panel presents the average number of friends in our dataset, while the Simulation Results panel presents the simulated number of friends.

Table A.9: The Difference in the Number of Friends (White - Black) Before and After Composition Changes

Panel A: High-Performing Friends			
	GPA 1-2	GPA 2-3	GPA 3-4
Simulation	0.31	0.75	1.05
Counterfactual 1: Equal Share	0.17	0.21	0.16
Counterfactual 2: Equal Share & Equal GPA Dist.	0.16	0.11	0.07
Panel B: Low-Performing Friends			
	GPA 1-2	GPA 2-3	GPA 3-4
Simulation	0.08	0.01	-0.26
Counterfactual 1: Equal Share	0.01	-0.01	-0.01
Counterfactual 2: Equal Share & Equal GPA Dist.	0.03	-0.01	-0.15

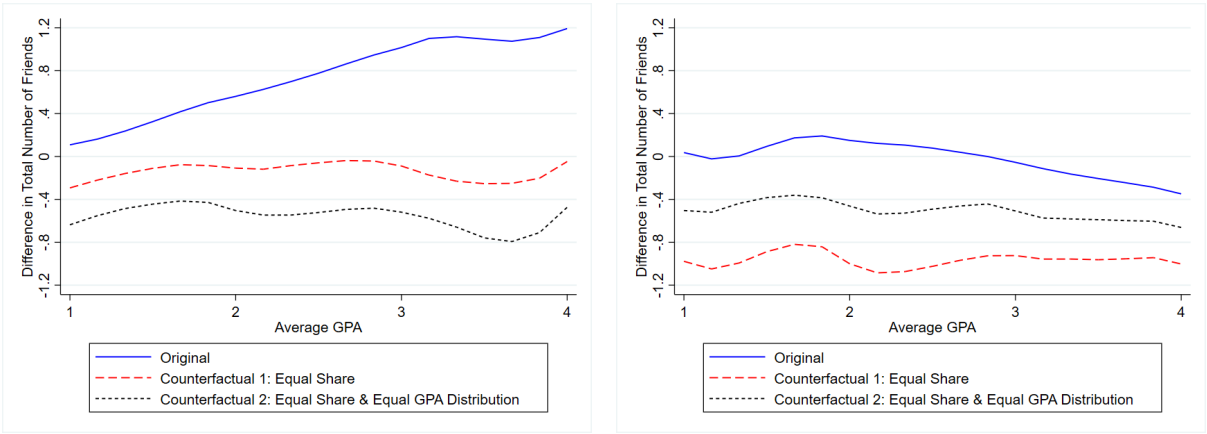
Notes. The table illustrates the difference in the number of expected friends between Black and White students, with standardization of 0 at the lowest GPA level. The expected numbers are computed using equation (3). Panel (a) and (b) present the difference in the number of high-performing and low-performing friends separately. In the table, the simulation presents the original difference in the expected number of friends. Counterfactual 1 presents the difference assuming an equal proportion of Black and White students. Counterfactual 2 presents the difference assuming both an equal share of Black and White students and an identical GPA distribution for Black and White students.

Fig. A.5. Difference in Number of Simulated Friends by GPA: White - Black



Notes. The figure illustrates the difference in the number of expected friends between Black and White students, standardized to 0 at the lowest GPA level following equation (3). In the graph, the solid blue line depicts the original difference in the expected number of friends. The dashed red line represents the difference assuming an equal proportion of Black and White students. The short-dashed black line represents the difference assuming both an equal share of Black and White students and an identical GPA distribution for Black and White students.

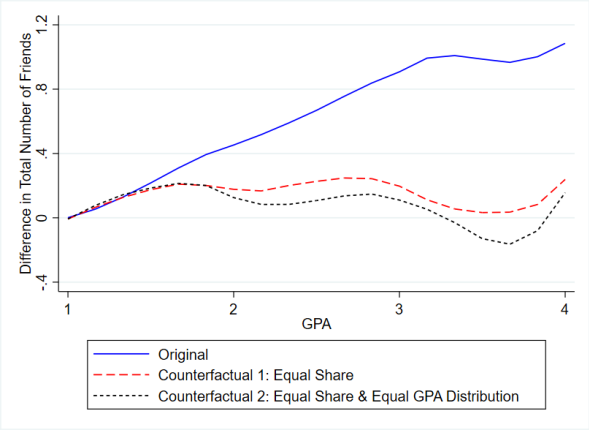
Fig. A.6. Difference in the Number of Simulated Friends by GPA Without Normalization: White - Black



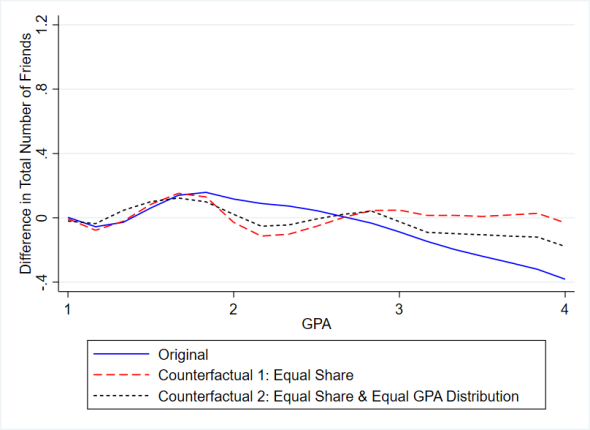
(a) High achieving Friends **(b)** Low achieving Friends

Notes. The figure illustrates the difference in the number of expected friends between Black and White students. The simulated numbers are computed using equations (3). Sub-figure (a) and (b) present the difference in the number of high-achieving and low-achieving friends separately. In the graph, the solid blue line depicts the original difference in the expected number of friends. The dashed red line represents the difference assuming an equal proportion of Black and White students. The short-dashed black line represents the difference assuming both an equal share of Black and White students and an identical GPA distribution for Black and White students.

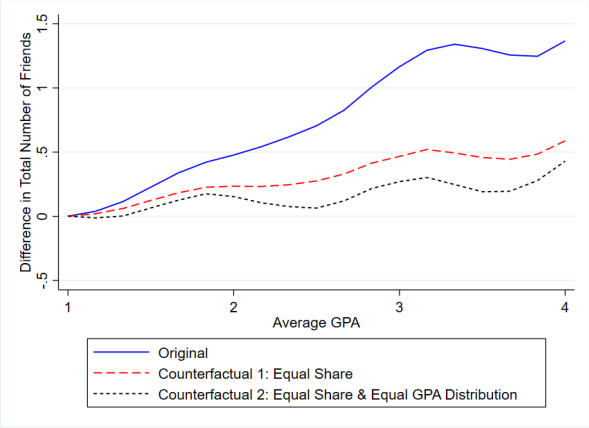
Fig. A.8. Racial Difference (White - Black) in the Number of Simulated Friends by GPA with Coefficients from Representative Schools



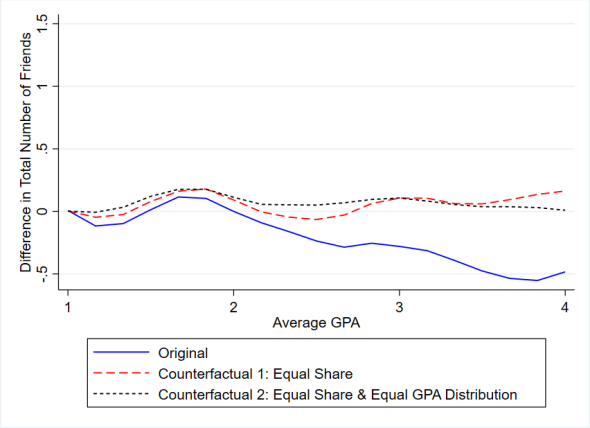
(a) All: High achieving Friends



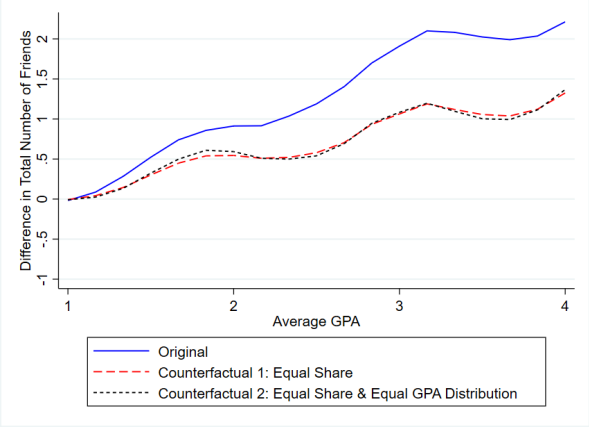
(b) All: Low achieving Friends



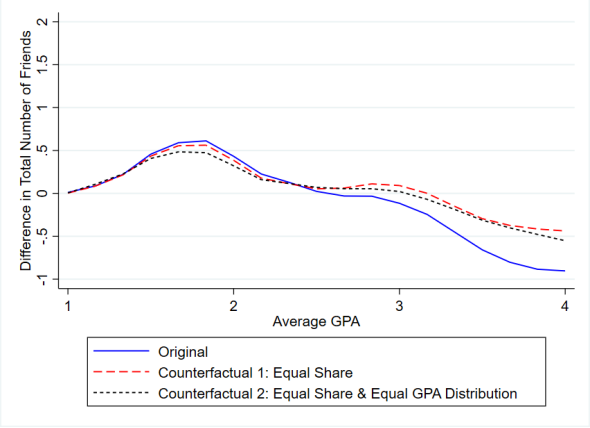
(c) Black 23%-40%: High achieving Friends



(d) Black 23%-40%: Low achieving Friends



(e) Black 41%-73%: High achieving Friends



(f) Black 41%-73%: Low achieving Friends

Notes. The figure illustrates the difference in the number of expected friends between Black and White students, standardized to 0 at the lowest GPA level following equation (3). Specifically, we use coefficients from the representative schools calculated in Table 2 and A.4. Sub-figure (a) and (b) use coefficients from all schools, Sub-figure (c) and (d) use coefficients from schools with Black enrollment between 23-40%, and Sub-figure (e) and (f) use coefficients from schools with Black enrollment between 41-74%. In the graph, the solid blue line depicts the original difference in the expected number of friends. The dashed red line represents the difference assuming an equal proportion of Black and White students. The short-dashed black line represents the difference assuming both an equal share of Black and White students and an identical GPA distribution for Black and White students.

A.7 Additional Earnings Outcomes

We further restrict our sample for labor market analysis to include individuals who report measures of cognitive skills, social skills, education level, and earnings. Consistent with Lleras-Muney et al. (2020), cognitive skills are assessed using the Add Health Picture Vocabulary Test (AHPVT) score administered in Wave 1, while social skills are gauged through self-reported extroversion levels collected in Wave 2.[A.4] Earnings are measured as total earnings from wages or salary in the previous year. We utilize earnings data from Waves 4, corresponding to ages 18-26. Respondents who responded "do not know" to the earnings question were presented with twelve earnings categories, and we approximate their earnings by using the midpoint of the selected range. We drop individuals who reported zero earnings (7%). On average, individuals in our sample earned approximately $37,300 in the preceding year. See Appendix Table A.10 for summary statistics of the earnings sample.

Appendix Table A.11 presents the first-stage results of our IV estimation. In column (1), we use age distance to peers in the same school and grade as our instrument. The distance is the average of the absolute difference between all possible pairs in the group with a student. In column (2), we disaggregate the age distance into two components: distance to Black, White, and other than Black and White peers in the same school and grade. Thus, each student has three age-distance variables. The rationale is that the age distance measure may predominantly reflect distance to White students, even if Black students are more inclined to form friendships with other Black students. Moreover, in column (2) we interact the age distance with Black and other race student dummies to examine potential racial differences in the sensitivities to friendships. Students are more sensitive to age differences when they form friendships with their same-race peers. In column (3), we disaggregate the age distance into low- and high-achieving peers, and interact these with a low-achieving dummy to consider homophily in achievement. Again, we find stronger sensitivity to the same group. Columns (1), (2), and (3) are matched to columns (3) and (4) of Table 3, and column (6) of Appendix Table A.12, respectively. In all cases, we

[A.4] The survey question is “You are shy?”, and the choices are “strongly agree, agree, neither agree nor disagree, disagree, strongly disagree”. Individuals choosing last three categories are defined as extrovert. Due to the survey design, 26% of the individuals in the dataset lack this information. Consequently, we impute this measure and incorporate a dummy variable to indicate its absence.

observe a strong relationship between friendship and age distance variables.

In Appendix Table A.12, we present supplementary information to Table 3, including results for students of races other than Black and White, and differentiated returns for high- and low-achieving students. Students of races other than Black and White do not show any difference from White students in returns for one more friend. Column (3) presents estimation results for high- and low-achieving students separately. The results indicate that low-achieving students have smaller returns from additional friends, which might reflect differences in the friends they have due to homophily. While having one more friend correlates with a 2.9% increase in earnings for high-achieving students, it corresponds to a 1.5% increase for low-achieving students. This difference is also observed in the 2SLS results in column (6) while the estimation results are not precise. Again, the results remain within the estimated bounds from Lleras-Muney et al. (2020).

Table A.10: Summary, Individual Level (Earnings Sample)

	White	Black	Other	All
GPA	2.9	2.6	2.9	2.8
Female (%)	51	58	51	53
Age	16	16	16	16
Number of Friends	3.80	2.97	2.88	3.30
High Performing (GPA$\geq$3)	2.47	1.58	1.66	2.00
Low Performing (GPA$<$3)	1.33	1.40	1.22	1.31
Cognitive Skills	106	97	98	101
Social Skills	0.47	0.44	0.38	0.43
Share (%)	44	21	35	100
Observations	3,957	1,949	3,175	9,081

Table A.11: First-Stage: Effects of Age Distance on the Number of Friendships

Outcomes are Number of Friends	(1)	(2)	(3)
Age distance	-1.102***		
	(0.121)		
Age distance to Black Peers		0.068	
		(0.108)	
Age distance to White Peers		-0.574	
		(0.359)	
Age distance to Other Peers		-0.337	
		(0.340)	
Age distance to Black Peers× Black		-0.722*	
		(0.424)	
Age distance to White Peers× Black		0.134	
		(0.469)	
Age distance to Other Peers× Black		0.424	
		(0.458)	
Age distance to Black Peers× Other		-0.146	
		(0.153)	
Age distance to White Peers× Other		0.552	
		(0.437)	
Age distance to Other Peers× Other		-0.588	
		(0.498)	
Age distance to Low-Achieving Peers			-0.537
			(0.376)
Age distance to High-Achieving Peers			-0.397
			(0.388)
Age distance to Low-Achieving Peers× Low-Achieving			-0.317
			(0.516)
Age distance to High-Achieving Peers× Low-Achieving			-0.074
			(0.493)
F-stat of Homophily Measures	83.0	9.0	30.4
P-value	0.000	0.000	0.000
R^2	0.041	0.041	0.046
Observations	9,059	9,059	9,047

Notes. In the table, “Controls" stands for the inclusion of control variables, including age, sex, race, social skill measure, cognitive skill measure, and school and grade fixed effects. Age distance is the average age distance to all students in the same school and grade. Age distance to Black (white) peers is the average age distance to all Black (White) students in the same school and grade. Black (White) is the dummy variable taking 1 if a student is Black (White). We use students of all races. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

Table A.12: Effects of the Number of Friends on Earnings: Appendix

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Friends	0.024***		0.029***	0.146*	0.088	0.117
	(0.004)		(0.004)	(0.081)	(0.092)	(0.072)
Friends× Black	-0.002				-0.006	
	(0.012)				(0.069)	
Friends× Other	0.002				0.015	
	(0.006)				(0.013)	
Friends× Low-Achieving			-0.014***			-0.019*
			(0.004)			(0.010)
High-Achieving Friends		0.029***				
		(0.004)				
Low-Achieving Friends		0.011*				
		(0.006)				
High-Achieving Friends× Black		0.003				
		(0.013)				
Low-Achieving Friends× Black		-0.005				
		(0.014)				
High-Achieving Friends× Other		-0.004				
		(0.007)				
Low-Achieving Friends× Other		0.014				
		(0.009)				
Controls	X	X	X	X	X	X
IV				Overall	Racial	GPA
Observations	9059	9059	9059	9059	9059	9047

Notes. The table presents estimation results from equation (5). Columns (1-3) present OLS estimates and columns (3-6) present 2SLS estimates in which the number of friends is endogenous. See A.7 for first-stage results. “Controls" stands for the inclusion of control variables, including age, sex, race, social skills, cognitive skills, and school and grade fixed effects. “Overall” indicates average age distance for all students in a school-grade. “Race-Specific" indicates the race-specific average age distance in school-grade. “Achievement-Specific" indicates the achievement-specific, high- and low-achieving, age distance in school-grade. “Black” and “Other” indicate a dummy variable taking the value of 1 if the race is Black and other than Black and White, respectively. We use students of all races. To facilitate the interpretation of the “Friends" variables, we include an interaction term between the number of friends (or high- and low-achieving friends) and a dummy variable for races other than Black and White. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

B Robustness Checks

B.1 Mutual Friendship Measures

Table B.1: Friendship Formation by Race and Achievement: Mutual Friendship (Both)

	Dependent variable : *i* nominating *j* as a friend (%)				
	(1)	(2)	(3)	(4)	(5)
dGPA	-0.095***	-0.078***	-0.112***	-0.093***	-0.131***
	(0.010)	(0.008)	(0.013)	(0.011)	(0.015)
$dGPA \times BB$			0.036*	0.022	0.034
			(0.021)	(0.022)	(0.025)
$dGPA \times BW$			0.100***	0.086***	0.104***
			(0.013)	(0.012)	(0.013)
$dGPA \times WB$			0.100***	0.086***	0.104***
			(0.013)	(0.012)	(0.013)
BB			0.002	0.026	0.928***
			(0.065)	(0.069)	(0.108)
BW			-0.295***	-0.296***	0.000
			(0.035)	(0.029)	(0.000)
WB			-0.295***	-0.296***	0.000
			(0.035)	(0.029)	(0.000)
Constant	0.289***	0.088	0.335***	0.178***	0.373***
	(0.026)	(0.054)	(0.036)	(0.064)	(0.021)
Observations	36,183,256	36,183,256	36,183,256	36,183,256	36,183,256
Mean	0.208	0.208	0.208	0.208	0.208
Controls		X		X	X
Individual FE					X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student *i* and student *j* both listed each other as a friend. *dGPA* indicates a difference in average GPA within a pair of students. *B* indicates a Black student, and *W* indicates a White student. *BW* is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

In our dyadic analysis, we measure friendships based on directed numbers, where one student (i) directs another (j) as their friend. However, in Appendix Table B.1, we examine the sensitivity of our friendship measure by using a stricter definition of friendship: mutual friendship, where both students mutually direct each other as friends. While this stricter definition results in a smaller probability of forming friendships compared to the baseline measure, our findings are

qualitatively unchanged. The difference in GPA matters again similarly between pairs of Black students and pairs of White students ($dGPA \times BB$). Additionally, we find that the difference in GPA matters less on across-race friendship formations compared to within-race friendships ($dGPA \times BW$ and $dGPA \times WB$).

Table B.2: Friendship Formation by Race and Achievement: Mutual Friendship (Either)

	Dependent variable : i nominating j as a friend (%)				
	(1)	(2)	(3)	(4)	(5)
$dGPA$	-0.282***	-0.251***	-0.319***	-0.284***	-0.388***
	(0.028)	(0.023)	(0.037)	(0.032)	(0.041)
$dGPA \times BB$			0.058	0.012	0.041
			(0.067)	(0.069)	(0.074)
$dGPA \times BW$			0.270***	0.229***	0.289***
			(0.036)	(0.033)	(0.038)
$dGPA \times WB$			0.270***	0.229***	0.289***
			(0.036)	(0.033)	(0.038)
BB			0.334	0.475*	3.574***
			(0.267)	(0.275)	(0.398)
BW			-0.926***	-0.935***	0.000
			(0.122)	(0.093)	(0.000)
WB			-0.926***	-0.935***	0.000
			(0.122)	(0.093)	(0.000)
Constant	1.060***	1.299***	1.168***	1.546***	1.503***
	(0.094)	(0.147)	(0.125)	(0.200)	(0.080)
Observations	36,183,256	36,183,256	36,183,256	36,183,256	36,183,256
Mean	0.817	0.817	0.817	0.817	0.817
Controls		X		X	X
Individual FE					X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes on the value of 1 if at least one of students i and j listed another student as a friend. $dGPA$ indicates a difference in average GPA within a pair of students. B indicates a Black student, and W indicates a White student. BW is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

Moreover, we alternatively define the mutual friendship measure, where if at least one student directs the other as a friend, both the $i - j$ and $j - i$ pairs are considered as forming friendships. The results are presented in Appendix Table B.2. This looser definition of friendship results in a higher probability of forming friendships compared to our baseline measure. However,

our findings are qualitatively unchanged. The difference in GPA matters once again similarly between pairs of Black students and pairs of White students. We find that the difference in GPA matters less on across-race friendship formations compared to within-race friendships.

B.2 Asymmetric GPA Measures

In Appendix Table B.3, we further examine whether the GPA gap matters differently when students direct another with a higher or lower GPA. The variable *ReceivingHighGPA* indicates that the sender student (i) has a lower GPA than the receiver student (j). We find that the relationship between the GPA gap and friendship formation is weaker when the receiver has a higher GPA ($ReceiverHighGPA \times dGPA$). This suggests that the difference in GPA matters differently when a student directs another student with a higher or lower GPA. However, we do not find a significant difference between White-White pairs and Black-Black pairs, regardless of the direction from lower GPA to higher GPA or higher GPA to lower GPA ($ReceiverHighGPA \times dGPA \times BB$ and $dGPA \times BB$).

Table B.3: Friendship Formation by Race and Achievement: Asymmetric GPA

	Dependent variable : i nominating j as a friend (%)	
	(1)	(2)
$ReceiverHighGPA \times dGPA$	0.032***	0.132***
	(0.006)	(0.016)
$ReceiverHighGPA \times dGPA \times BB$	-0.018	-0.015
	(0.021)	(0.020)
$ReceiverHighGPA \times dGPA \times BW$	-0.040***	-0.046***
	(0.009)	(0.014)
$ReceiverHighGPA \times dGPA \times WB$	0.003	0.006
	(0.012)	(0.014)
$ReceiverHighGPA \times BB$	0.033	0.029
	(0.025)	(0.023)
$ReceiverHighGPA \times BW$	0.021**	0.017*
	(0.008)	(0.010)
$ReceiverHighGPA \times WB$	0.024**	0.031***
	(0.011)	(0.011)
$ReceiverHighGPA$	-0.030***	-0.019***
	(0.007)	(0.005)
$dGPA$	-0.229***	-0.253***
	(0.027)	(0.028)
$dGPA \times BB$	0.054	0.022
	(0.044)	(0.045)
$dGPA \times BW$	0.209***	0.187***
	(0.026)	(0.024)
$dGPA \times WB$	0.186***	0.158***
	(0.026)	(0.025)
BB	0.153	0.240
	(0.164)	(0.169)
BW	-0.631***	-0.636***
	(0.079)	(0.061)
WB	-0.608***	-0.616***
	(0.079)	(0.061)
Constant	0.762***	0.868***
	(0.081)	(0.129)
Observations	36,519,842	36,183,256
Mean	0.5	0.5
Controls		X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominates student j as a friend. *ReceiverHighGPA* takes the value of 1 if the nominated student j has higher GPA than the student i. *dGPA* indicates a difference in average GPA within a pair of students. B indicates a Black student, and W indicates a White student. BW is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

B.3 Semiparametric GPA Measures

We explore whether low and high GPA students exhibit different patterns when they form friendships with low and high GPA peers. We divide students into two GPA groups: low (average GPA 1-3) and high (average GPA 3-4). Then, we examine the difference across races when a student i with a low (or high) GPA forms a friendship with a student j with a low (or high) GPA. The results are presented in Appendix Table B.4. First, we observe a higher probability of forming friendships within the same GPA groups than across GPA groups. Particularly, stronger homophily exists among high GPA students. Regardless of whether a student with a low (or high) GPA forms friendships with a student with a low (or high) GPA, however, the GPA gap matters less or similarly to Black-Black pairs compared to White-White pairs. Additionally, we further explore the differences using three levels of GPA: low (average GPA 1-2.5), middle (average GPA 2.5-3.25), and high (average GPA 3.25-4), as shown in Appendix Table B.5. Once again, we find that the GPA gap matters similarly or less on friendships for Black-Black pairs compared to White-White pairs.

Table B.4: Friendship Formation by Race and Achievement: Semiparametric Low-High

	Dependent variable : i nominating j as a friend (%)	
	(1)	(2)
$GPAi : low \times GPAj : high \times BW$	0.060***	0.053**
	(0.022)	(0.021)
$GPAi : low \times GPAj : high \times WB$	0.113***	0.104***
	(0.025)	(0.021)
$GPAi : low \times GPAj : high \times BB$	0.076	0.069**
	(0.047)	(0.031)
$GPAi : high \times GPAj : low \times BW$	0.115***	0.106***
	(0.024)	(0.021)
$GPAi : high \times GPAj : low \times WB$	0.060**	0.053**
	(0.023)	(0.023)
$GPAi : high \times GPAj : low \times BB$	0.040	0.032
	(0.045)	(0.027)
$GPAi : high \times GPAj : high \times BW$	-0.214***	-0.221***
	(0.075)	(0.055)
$GPAi : high \times GPAj : high \times WB$	-0.209***	-0.216***
	(0.075)	(0.053)
$GPAi : high \times GPAj : high \times BB$	0.174	0.126
	(0.244)	(0.192)
$GPAi : low \times GPAj : high$	-0.074***	-0.066***
	(0.021)	(0.009)
$GPAi : high \times GPAj : low$	-0.087***	-0.089***
	(0.023)	(0.012)
$GPAi : high \times GPAj : high$	0.298***	0.283***
	(0.075)	(0.041)
BB	0.214*	0.237*
	(0.114)	(0.129)
BW	-0.452***	-0.499***
	(0.057)	(0.052)
WB	-0.429***	-0.473***
	(0.057)	(0.052)
Constant	0.548***	0.846***
	(0.058)	(0.124)
Observations	36,183,256	36,183,256
Mean	0.5	0.5
Controls	X	X

Notes. The table presents coefficients and standard errors from estimation of equation (1), with GPA difference defined in intervals (high and low) rather than continuously. Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominates student j as a friend. B indicates a Black student, and W indicates a White student. BW is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects. *** $p < .01$, ** $p < .05$, * $p < .1$

Table B.5: Forming Friends and Difference in GPA: Semiparametric Low-Middle-High

	Dependent variable : i nominating j as a friend (%)	
	(1)	(2)
$GPAi : low \times GPAj : middle \times BW$	0.086***	0.093***
	(0.018)	(0.024)
$GPAi : low \times GPAj : middle \times WB$	0.093***	0.081***
	(0.017)	(0.020)
$GPAi : low \times GPAj : middle \times BB$	0.046	0.057
	(0.044)	(0.039)
$GPAi : low \times GPAj : high \times BW$	0.187***	0.167***
	(0.032)	(0.034)
$GPAi : low \times GPAj : high \times WB$	0.241***	0.217***
	(0.035)	(0.034)
$GPAi : low \times GPAj : high \times BB$	0.107**	0.099**
	(0.052)	(0.039)
$GPAi : middle \times GPAj : low \times BW$	0.104***	0.090***
	(0.018)	(0.024)
$GPAi : middle \times GPAj : low \times WB$	0.080***	0.086***
	(0.020)	(0.026)
$GPAi : middle \times GPAj : low \times BB$	0.076*	0.084***
	(0.041)	(0.031)
$GPAi : middle \times GPAj : middle \times BW$	0.003	0.021
	(0.031)	(0.039)
$GPAi : middle \times GPAj : middle \times WB$	-0.001	0.018
	(0.033)	(0.038)
$GPAi : middle \times GPAj : middle \times BB$	0.076	0.094
	(0.097)	(0.078)
$GPAi : middle \times GPAj : high \times BW$	0.012	-0.008
	(0.047)	(0.043)
$GPAi : middle \times GPAj : high \times WB$	0.032	0.041
	(0.048)	(0.042)
$GPAi : middle \times GPAj : high \times BB$	0.230	0.204
	(0.173)	(0.139)
$GPAi : high \times GPAj : low \times BW$	0.248***	0.224***
	(0.033)	(0.033)
$GPAi : high \times GPAj : low \times WB$	0.179***	0.160***
	(0.033)	(0.035)
$GPAi : high \times GPAj : low \times BB$	0.090*	0.083**
	(0.051)	(0.038)
$GPAi : high \times GPAj : middle \times BW$	0.068	0.078*
	(0.047)	(0.040)
$GPAi : high \times GPAj : middle \times WB$	0.019	0.001
	(0.047)	(0.043)
$GPAi : high \times GPAj : middle \times BB$	0.153	0.131
	(0.151)	(0.117)

Table B.5: Friendship Formation by Race and Achievement: Semiparametric Low-Middle-High (Continued)

	Dependent variable : i nominating j as a friend (%)	
	(1)	(2)
$GPAi : high \times GPAj : high \times BW$	-0.207**	-0.229***
	(0.091)	(0.070)
$GPAi : high \times GPAj : high \times WB$	-0.209**	-0.229***
	(0.090)	(0.066)
$GPAi : high \times GPAj : high \times BB$	0.233	0.165
	(0.305)	(0.245)
$GPAi : low \times GPAj : middle$	-0.085***	-0.067***
	(0.015)	(0.009)
$GPAi : low \times GPAj : high$	-0.211***	-0.178***
	(0.030)	(0.020)
$GPAi : middle \times GPAj : low$	-0.098***	-0.087***
	(0.017)	(0.011)
$GPAi : middle \times GPAj : middle$	0.018	0.031**
	(0.028)	(0.015)
$GPAi : middle \times GPAj : high$	0.013	0.037**
	(0.045)	(0.016)
$GPAi : high \times GPAj : low$	-0.229***	-0.210***
	(0.032)	(0.024)
$GPAi : high \times GPAj : middle$	-0.005	0.014
	(0.044)	(0.015)
$GPAi : high \times GPAj : high$	0.307***	0.318***
	(0.090)	(0.050)
BB	0.173*	0.197*
	(0.101)	(0.118)
BW	-0.500***	-0.544***
	(0.061)	(0.055)
WB	-0.469***	-0.511***
	(0.061)	(0.055)
Constant	0.592***	0.877***
	(0.062)	(0.124)
Observations	36,183,256	36,183,256
Mean	0.5	0.5
Controls		X

Notes. The table presents coefficients and standard errors from estimation of equation (1), with GPA difference defined in intervals (high, middle, and low) rather than continuously. Units of observation are directed dyads, based on the non-Hispanic Black and non-Hispanic White students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominates student j as a friend. B indicates a Black student, and W indicates a White student. BW is an indicator that the sender is a Black student and the receiver is a White student. Mean is the average probability of forming friendships. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects. Individual FE indicates that individual fixed effects of senders and receivers are controlled for. Standard errors are clustered at the school level. *** $p < .01$, ** $p < .05$, * $p < .1$

B.4 Including Hispanic Students

We examine whether including Hispanic students affects our results. We add Hispanic students to our analysis sample and additional variables indicating Hispanic students to the regression equation (1). The added variables are as follows: B_iH_j, W_iH_j, H_iH_j, H_iW_j, H_iB_j, $d_{ij} \times B_iH_j$, $d_{ij} \times W_iH_j$, $d_{ij} \times H_iH_j$, $d_{ij} \times H_iW_j$, and $d_{ij} \times H_iB_j$. As presented in Appendix Table B.6, including Hispanic students does not significantly affect the results. The difference in GPA matters similarly between pairs of Black students and pairs of White students ($dGPA \times BB$) and matters less between pairs of Black and White students ($dGPA \times BW$).

Table B.6: Friendship Formation by Race and Achievement: Including Hispanic

	Dependent variable : i nominating j as a friend (%)	
	(1)	(2)
$dGPA$	-0.216***	-0.195***
	(0.025)	(0.022)
$dGPA \times BB$	0.047	0.022
	(0.044)	(0.045)
$dGPA \times BW$	0.188***	0.163***
	(0.025)	(0.023)
$dGPA \times BH$	0.186***	0.158***
	(0.028)	(0.026)
$dGPA \times WB$	0.182***	0.160***
	(0.024)	(0.023)
$dGPA \times WH$	0.110***	0.104***
	(0.023)	(0.020)
$dGPA \times HH$	0.151***	0.117***
	(0.034)	(0.034)
$dGPA \times HW$	0.117***	0.107***
	(0.023)	(0.019)
$dGPA \times HB$	0.180***	0.152***
	(0.028)	(0.025)
BB	0.168	0.234
	(0.166)	(0.152)
BW	-0.624***	-0.612***
	(0.078)	(0.059)
BH	-0.541***	-0.324***
	(0.095)	(0.066)
WB	-0.597***	-0.587***
	(0.078)	(0.058)
WH	-0.360***	-0.261***
	(0.071)	(0.050)
HH	-0.427***	-0.060
	(0.132)	(0.074)
HW	-0.383***	-0.281***
	(0.070)	(0.048)
HB	-0.535***	-0.315***
	(0.094)	(0.067)
Constant	0.752***	0.876***
	(0.080)	(0.092)
Observations	53,008,910	53,008,910
Mean	0.433	0.433
Controls		X

Notes. The table presents coefficients and standard errors from estimation of equation (1). Units of observation are directed dyads, based on the non-Hispanic Black, non-Hispanic White, and Hispanic students who reported GPAs and individual characteristics (age, gender, and race). The dependent variable takes the value of 1 if student i nominates student j as a friend. $dGPA$ indicates a difference in average GPA within a pair of students. B, W, and H indicates a Black, White, and Hispanic student, respectively. Controls indicate the controlling demographic variables including gender and age of senders and receivers, difference in age within pairs, and school fixed effects.
*** $p < .01$, ** $p < .05$, * $p < .1$