NBER WORKING PAPER SERIES

THE DETERMINANTS AND CONSEQUENCES OF FRIENDSHIP COMPOSITION

Jason M. Fletcher
Stephen L. Ross
Yuxiu Zhang

Working Paper 19215
http://www.nber.org/papers/w19215

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2013

The authors thank the NICHD for financial support. The authors benefited from helpful suggestions by participants at seminars at University of Connecticut, University at Albany-SUNY, Federal Reserve Bank of Atlanta, the NBER's Education Program and Childrens Program Joint Meetings, and the 2012 Add Health Users Conference. We thank Edward Vytlacil, Ken Frank, Janet Currie, Caroline M. Hoxby, Kevin Lang, Damon Clark and David Frisvold for valuable comments. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number R21HD066230. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Bureau of Economic Research. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w19215.ack

The Determinants and Consequences of Friendship Composition
Jason M. Fletcher, Stephen L. Ross, and Yuxiu Zhang
NBER Working Paper No. 19215
July 2013
JEL No. I20,I21,J24

ABSTRACT

This paper examines the demographic pattern of friendship links among youth and the impact of those patterns on own educational outcomes using the friendship network data in the Add Health. We develop and estimate a reduced form matching model to predict friendship link formation and identify the parameters based on across-cohort, within school variation in the "supply" of potential friends. Our model provides novel evidence on the impact of small changes in peer demographic composition on the pattern of friendship links. The evidence suggests, for example, that increases in the share of African-American or Hispanic students leads to reductions in the incidence of cross race friendships. We then use the predicted friendship links from the model in an instrumental variable analysis of the effects of friends' socioeconomic status, as measured by parental education, on own grade point average outcomes. Although the conditional correlation between friendship composition and grade point average suggests large associations between friends' characteristics and own grades, this effect is robust only for females in the instrumental variable analysis. We then present evidence that the GPA effects are driven by science and English grades and a mechanism is likely through non-cognitive factors.

Jason M. Fletcher
Yale School of Public Health
Department of Health Policy and Management
60 College Street, #303
New Haven, CT 06520
and NBER
jason.fletcher@yale.edu

Stephen L. Ross
University of Connecticut
Department of Economics
341 Mansfield Road, Unit 1063
Storrs, CT 06269-1063
stephen.l.ross@uconn.edu

Yuxiu Zhang
Yale School of Public Health
60 College St
New Haven, CT 06510
yuxiu909@gmail.com

Introduction

A growing body of evidence has documented the effects of peers on the academic outcomes of school children.[1] The effect of peers on such outcomes raises the natural policy question of what would really happen if peers were changed. A change in peer composition may alter the social dynamics of a school or other social network, and as a result alter the social ties and personal interactions through which peer effects may operate. Several recent studies (Fletcher and Ross 2012, Calvó-Armengol et al. 2009, Lavy and Sand 2013) document the effect of friendship and social networks on student and youth outcomes. Weinberg (2007) shows that students tend to associate with individuals like themselves, which may mitigate the impact of any change in peer composition. Mayer and Puller (2008) show evidence that increasing the opportunities of heterophilous (e.g. cross-race) relationships is not sufficient to substantially increase these links. Finally, Carrell, Sacerdote and West (2011) conduct a policy experiment where Air Force Academy students are assigned to work groups intended to maximize the performance of the lowest ability students. They find that in their treatment group, students sort into subgroups based on ability, eliminating the positive peer effects identified in earlier studies of the same environment, indicating the need for further information on how friendships form before policies can be suggested. However, the difficulties with estimating the friendship matching process in "real world" settings are numerous. This paper combines a quasi-experimental research design within the setting of high school friendship network formation to extend the literatures on friendship formation, as well as estimating the consequences of friend composition on educational outcomes.

In particular, this study uses the friendship nomination data in the Add Health sample to study the causal impact of small changes in peer composition on the demographic pattern of friendship formation.[2] We focus on within grade (or cohort) friendships, which represent a supermajority (83% of same-sex nominations[3]) of friendship ties in our sample, and identify the effect of peer composition on friendship formation by exploiting across cohort and within school variation in the composition of students (i.e. "potential friends"). Specifically, we classify potential friendship ties within grade based on a demographic match between each pair of students, and examine heterogeneity in the effects of cohort demographic composition on the likelihood of friendship formation conditional on demographic type of student pair by school fixed effects. Balancing tests confirm that cohort composition is orthogonal to incidental student attributes within school-pair type cells. Across cohort variation is regularly exploited in studies of the effects of peers on student outcomes (beginning with Hoxby 2000), but to our knowledge this is the first study to exploit this variation in order to examine friendship formation.[4] Our

---

[1] See Lavy and Schlosser (2011), Bifulco, Fletcher and Ross (2011) and references contained.

[2] Few other national datasets contains information on nominated best friends. Additionally many datasets contain a single grade-level (cohort) from each sampled school (e.g. NELS, ECLS-K, ELS, etc).

[3] Our focus is within school friendship. Among all within school same-sex friend nominations in which both parties have identifiable student id, school id and grade id, 17% are cross-grade nominations. Among all same-sex nominations with identifiable friend id (including those with missing school/grade id), 66% are same-grade within school; 14% are cross-grade within school; for another 19% nominations, nominated friends' id is identifiable, but school id and grade id are not, therefore we don't know whether they are within school/grade or not; for the rest 1%, we know the two parties in the nominations are from two different identifiable schools. In terms of out-of-school nomination, we need to take account unidentifiable nominations coded as 77777777 or 88888888. The proportion of out-of-school nominations among all is about 15%, and the proportion of within school nominations with unidentifiable student id is 7%.

[4] Perhaps the closest papers to our study in this regard are those by Fisman and colleagues (2008), who used random assignment during speed dating interactions to estimate the preferences for same and opposite-race social ties (i.e.

model of friendship formation focuses on peer maternal education based on the importance of parental education for child outcomes (Haveman and Wolfe 1995) and a previous finding in the same sample that peer's maternal education has a significant impact of academic outcomes (Bifulco, Fletcher and Ross 2011). We also examine race, ethnicity and gender, given the well-known concentration of friendships among students of the same race, ethnic and gender (Moody 2001). Student race/ethnicity is especially important to investigate given the large race/ethnic differences in educational attainment in the population.

Specifically, we examine how differences in the socio-demographic composition of a cohort relative to the composition of the other cohorts in the same school affect the likelihood of any pair of same gender students to mutually identify each other as friends. On maternal education, our key sorting results are for females, and we find that as the number of students whose mothers have a college degree increases relative to mothers with high school degrees, friendships between female students whose mothers both have a college degree or where one mother has a college degree and the other completed high school become more likely. Our estimates imply that one standard deviation increase in the share of maternal college students scaled by cohort size by gender is on average associated with a 23 percent and 36 percent increase in the probability of forming links between two maternal college students and between a maternal college and a maternal high school graduate student, respectively. Given the focus on friendship link formation, these results cannot be driven simply by an increase in the opportunity for college educated friends, but instead are consistent with an increase in the attractiveness of maternal college educated friends as the number of maternal college educated students rises. On race/ethnicity, differences in the share black have the largest impact, with white-white friendships becoming more frequent for both men and women and Other-Other (predominantly Asian-Asian) friendships becoming more frequent for women. Hispanic-Other friendships also become less likely among women. Increases in share Hispanic lead to lower rates of female across race friendships (white-Hispanic and Hispanic-Other) and higher rates of black-black friendships among men. Therefore, increases in minority representation appear to increase the level of homophily in friendship formation. We know of no other work that documents this shift towards homophily as the population of minority groups increases.[5]

We next examine the effect of friendship patterns on student outcomes. Building on earlier work (Bifulco, Fletcher and Ross 2011), we focus our analysis on the impact of friends' maternal education levels on academic outcomes. As discussed by Manski (1993, 1995, 2000) and others, research of the effects of social interactions between individuals must address several empirical issues because individuals select into friendships and peer groups. In order to address these concerns, we use our estimated model of the formation of friendship links in order to develop predictions of (i.e. instruments for) friendship composition for individual students. These instruments are highly predictive of individual student's actual friendship patterns even though the predictions do not contain any information on the individual's friendship patterns and are only identified by across-cohort variation in the demographic composition of schools. We find that the number of friends with a college educated mother has a large positive effect on the

---

dates). As in that study, our study examines the effects of presumably exogenous changes in the opportunity set for forming interpersonal relationships

[5] This finding is consistent with Mayer and Puller (2008) and Weinberg (2007) among others. Further, the difficulty of producing heterophilous ties in groups when adding diversity is also a likely mechanism for why the experiment conducted by Carrell et al. (2013) that increased academic diversity among military squadrons reduced the outcomes of these individuals.

grade point average of female students, where a one standard deviation increase in the number of maternal college friends is associated with a 0.178 standard deviation increase in GPA. These effects operate primarily through higher grades in English and Science courses. Mechanism analyses suggest that having more maternal college friends improves students' assessment of themselves and their educational environment, and might also improve their mental health and reduce troublesome behaviors, which are all closely correlated with GPA. The mechanism analysis also shows that the majority of the effect of maternal college friends operates through an effect on students whose mothers have a college education for both GPA and the mechanism variables. While there is a strong conditional correlation between the number of friends with a college educated mother and grade point average for male students, these effects do not persist in our instrumental variable estimates. The effect of the number of high school drop-out friends is zero in both the OLS and IV estimates for both male and female students, indicating asymmetric effects of maternal schooling.

In terms of identification, the strength of the GPA effects of friends arises from the exogeneity of the instrument, and to our knowledge this is the first study of friendship effects that exploits such across cohort variation. The student level GPA regressions include school by student demographic type fixed effects. Further, similar to "leave-one-out" strategies used in Angrist, Imbens and Krueger (1999) and Blomquist and Dahlberg (1999), we address concerns about incidental parameters bias in the match model fixed effect estimates by calculating individual specific friendship predictions using individual specific fixed effects that omit any information associated with that individual's friendship choices. Following Guryan, Kroft and Notowidigdo (2009), we address the bias caused by omitting this information on the individual's choices by developing a school level control function for inclusion in the GPA regressions, and balancing tests confirm that the resulting instruments are not correlated with predetermined attributes conditional on the fixed effects and the control function. If there is a weakness of our identification strategy, it relates to our exclusion restrictions. While we rule out general cohort level peer effects by including grade by school fixed effects, we cannot rule out the possibility that a randomly assigned peer environment that leads to more friendships with students whose mothers have a college education for a given type student also directly increases girls' GPAs. In order to explore this concern, we examine whether the friendship match model for females can explain academic outcomes for males, and find that it cannot. The match model only predicts GPA for the subsample where it predicts friendships. Nonetheless, a somewhat weaker conclusion based on our results is that peer environments that raise the likelihood of a particular student having friends with college educated mothers leads to an increase in girls' GPAs.

Finally, for girls, we conduct detailed calculations examining the effect of an increase in the number of students with a mother who is a college graduate. The calculations examine the direct effect of adding more maternal college students both through the increase in the opportunity to form such friendships and the estimated effect of the share maternal college on the likelihood of friendship formation. A ten percentage point increase in the share maternal college in each cohort is associated with a 45 percent increase in the number of friends for the maternal college subsample, a 106 percent increase for the maternal high school graduate subsample, and a 109 percent increase for the maternal drop-out subsample. Most of these changes are associated with the increase in opportunities for friendships with maternal college students as opposed to the 9 and 14 percent effects of changing the probability of friendship formation.

Empirical Model of Friendship Formation

Consider a sample of schools (s) with a set of grades or cohorts (c) in each school. Students of a given gender may be systematically allocated to a school through their parents' choices, but are assumed to be distributed randomly across the cohorts or grades in any school because parents cannot easily observe the composition of individual cohorts when choosing a school, especially when those grade compositions will only be determined at a future time.[6]

Within a grade, every student can potentially form a friendship with any other student in the grade, and our student friendship data can be rearranged as a sample of pairs of students *i* and *j*. If a cohort has n_{cs} students of a given gender, then that population of students will contribute $n_{cs}(n_{cs}-1)/2$ total observations to our sample of potential friendship links. Pairs of students are categorized into one of *t* nominal "types" where a pair type is defined based on the demographic attributes of both students in the pair. Both the pairs and the pair types are defined to be non-directional so that reversing the order of the students does not create an additional observation or pair type.

The establishment of a social link between any pair of students (P_{ijtcs}) is a binary outcome that may be described by the following linear probability model

$$P_{ijtcs} = \tau_{cs}\left(\beta_t Z_{cs} + \delta_{ts} + \varepsilon_{is} + \varepsilon_{js} + \theta\varepsilon_{is}\varepsilon_{js} + \mu_{ijtcs}\right) \qquad (1)$$

where Z_{cs} is a 1 by m vector measuring the demographic composition of each cohort over types, β_t captures our behavior of interest by allowing the likelihood of friendship formation for each pair type *t* to vary with the demographic composition of the cohort. We assume that individual *i*'s demographic type is *x*, and individual *j* is type *y*. The pair formed by *i* and *j* is assigned pair type *t*, which represents the combination *{x,y}*. δ_{ts} allows the effect of belonging to that pair type *t* on friendship formation to vary by school so that the estimates of β_t are identified by across cohort comparisons of friendship patterns within school and friendship type, student unobservables on the propensity to form friendships are captured by random effects ε_{is} and ε_{js}, and finally μ_{ijcs} is a stochastic return to the match between these particular students. The probability is also scaled by τ_{cs} in order to allow the probability of friendship formation overall to vary across cohorts so that probabilities of formation with a particular individual can fall with cohort size. τ_{cs} captures the fact that number of friendship links grows approximately linearly with the cohort population while the number of potential friends is approximately quadratic.

Consistent estimation requires that

$$\begin{aligned} &E[\varepsilon_{is}|\tau_{cs}, Z_{cs}, t] = 0 \\ &E\left[\varepsilon_{is}\varepsilon_{js}|\tau_{cs}, Z_{cs}, t\right] = 0 \\ &E\left[\mu_{ijcs}|\tau_{cs}, Z_{cs}, t\right] = 0 \end{aligned} \qquad (2)$$

We believe that the assumptions in equation (2) are reasonable given our earlier assumption of the random allocation of students of each demographic type to a particular cohort *c* within a

[6] This assumption is supported in our sample by balancing tests conducted in Bifulco, Fletcher and Ross (2011) and later in this paper, demonstrating that individual attributes of students are not correlated with within-school variation in cohort composition.

school *s*, and the construction of the sample to include all possible pairs of students in a grade. Our and the literature's concern about bias arises from the potential correlation between ε_{is} and school composition (Z_{cs}) based on students (or their parents) sorting systematically into schools based on the demographic composition of those schools potentially violating the first condition in equation (2). Further, this sorting likely varies with the students' demographic attributes so that the conditional distribution of ε_{is} within school is not constant across students of different types. However, by linearly conditioning on school (*s*) by student pair type *t* fixed effects, we condition out the effect of sorting into schools on the mean of the distribution of ε_{is} for each observable student pair type and, given quasi-random assignment to cohorts within schools, ε_{is} should be uncorrelated with the within school variation in cohort demographics.[7]

Given the first assumption in equation (2), the only possible mechanism for violating the second assumption is if ε_{is} and ε_{js} are correlated within school and cohort. Specifically,

$$\begin{aligned} E\left[\varepsilon_{is}\varepsilon_{js}|\tau_{cs}, Z_{cs}, t\right] = \\ Cov\left[\varepsilon_{is}, \varepsilon_{js}|\tau_{cs}, Z_{cs}, t\right] + E[\varepsilon_{is}|\tau_{cs}, Z_{cs}, t]E\left[\varepsilon_{js}|\tau_{cs}, Z_{cs}, t\right] \\ = Cov\left[\varepsilon_{is}, \varepsilon_{js}|\tau_{cs}, Z_{cs}, t\right] \end{aligned} \qquad (3)$$

However, our sample of pairs within cohort are constructed to include all possible pairs of students and so, with the assumption of no selection into cohorts within schools, the correlation or covariance must be zero. Finally, it is relatively standard to assume that the idiosyncratic error associated with the match between two individuals μ_{ijcs} is orthogonal to the observables.[8]

In the context of our specific problem and data, we next specify the details of the model that we will estimate. First, we note that, asymptotically, τ_{cs} must be inversely proportional to the number of potential friends in cohort (n_{cs}) because otherwise the actual number of friends will limit to either 0 or infinity as the cohort size becomes larger. As a result we approximate τ_{cs} with $1/n_{cs}$ and estimate δ_{xys} and β_{xy} using the following equation

$$P_{ijtcs} = \beta_t \left(\frac{Z_{cs}}{n_{cs}}\right) + \tilde{\delta}_{ts} + \tilde{\mu}_{ijtcs} \qquad (4)$$

Note that, at least to a first order approximation, the pair-type by school fixed effects ($\tilde{\delta}_{ts}$) can be estimated as a common set of parameters across cohorts within a school, $\tilde{\delta}_{ts} = \frac{\delta_{ts}}{\bar{n}_{cs}} \approx \frac{\delta_{ts}}{n_{cs}}$, because with a moderate size or larger school and quasi-random allocation of students to cohorts

[7] The above claim relies on the implicit assumption that the expectation of ε_{is} conditional on pair type *t* is zero otherwise random variation in cohort racial composition will lead to systematic changes in the average unobservables of the individuals in a type and cohort. However, this restriction is a standard assumption in virtually all reduced form studies including studies that exploit random assignment because one cannot randomly assign the attributes of the randomly assigned factors, e.g. peers or environmental circumstances, and our analysis captures the causal effect of more students of a given type in a cohort on friendship formation including the effect through unobservables that are systematically associated with that type.

[8] In principle, one might question whether students have correlated unobservables in the same cohort because some of them will end up in the same classroom or share similar interests. However, such phenomena do not lead to a conditional correlation within the population unless that likelihood varies systematically across cohorts in the same school. The effect of the average probability of sharing a class or an interest with another student on friendship link formation should be captured by the school-student pair type fixed effects, and after conditioning out that effect the only obvious source of correlation is sorting, which our assumptions rule out.

n_{cs} is relatively constant within a school (near the mean - $\bar{n}_{cs}$) and deviations in n_{cs} within school can be treated as exogenous.[9]

Data and Estimation of the Friendship Model

Data Description

In order to examine the determinants and achievement consequences of friendship ties during high school, we use the only available dataset with information on nominated friends from multiple grade-levels in a large number of schools, the National Longitudinal Study of Adolescent Health (Add Health). Add Health is a school based longitudinal study of health and education-related behaviors of adolescents with follow up through age 30. For this paper, we focus on the "In-School" data collection, which utilized a self-administered survey to more than 90,000 students in grades 7 through 12 during a class period at school between September 1994 and April 1995. The survey focused on collecting data on socio-demographic characteristics, family background, health status, risk behaviors, academic achievement, school factors, and friendship nominations. Specifically, each student respondent was asked to identify up to five male and five female friends that attended the same school (these nominations were later cross-referenced with school rosters). Based on the friendship nominations, social networks within each school can be constructed, allowing data links between friends' reported background characteristics and respondent's reported course grades in English, math, science, and history courses.

Of the over 90,000 students originally surveyed, there are several sample size reductions necessary to create our analysis sample. 178 individuals were dropped from the sample due to missing identification numbers[10]; another 2,637 are dropped because of missing grade, race, sex, mom's education, or missing the majority of their friendship information; we exclude 116 observations from small schools (less than 40 students in school or less than an average of 10 students per grade); we exclude the twin sample, which contains 2,492 students. This process gives us an empirical sample of 84,695 coming from 139 schools with school size between 44 and 2,418 students allocated across two to four grades or cohorts.

Like much previous work, we focus on same-gender friendships in our analysis. The primary reason for this choice is to separate "friends" from "romantic relationships". We also limit our analysis to examining links between individuals in the same grade level. As we describe in more detail below, this focus allows us to utilize an across-cohort research design[11]. We focus on directed ties, and mutual friendship in particular, meaning two students are considered as a pair of friends if they both nominated each other. We assume the influence from

---

[9] One concern with equation (1) arises from the heteroskedasticity associated with the linear probability model. With similar number of friends at the individual level regardless of cohort size, the matrix of P_{ijtcs} becomes very sparse for large schools with large numbers of students in each cohort and is much more dense for smaller schools. Equation (4) addresses this by decreasing the magnitude of the independent variable for large cohorts/schools where the frequencies of non-zero P_{ijtcs} are very low rather than requiring the effect of cohort composition to be the same in percentage point terms for link formation in allowing for a lower probability of link frequency for these sparse regions of the social link vector.

[10] These individuals were likely new students and not yet on the school roster

[11] Although the focus on same-grade nominations may appear constricting, we note that over 80% of all nominations we capture in the data are for individuals in the same grade. We also regressed our cohort variables (i.e. the 'supply' of types of friends in a cohort) on whether individuals nominate friends outside of their grade with school-gender fixed effect, and found no correlation

friends to be strongest in a relationship which both parties in the pair agree on the friendship. It is also worth noticing that in Add Health, though a student can nominate up to five same-gender friends, not many students appear constrained by this cap. The average numbers of identifiable same-gender friends nominated are 2.65 for male students and 3.10 for female students. The majority of the nominations are one-direction. Therefore, the number of mutual ties is low. On average, a male has 0.69 and a female has 1.10 mutual friends.[12] We begin by showing the basic friendship patterns in the data on our key variables of interest. Table 1 shows the fraction of same-gender/same-grade friendships in each maternal education category by the maternal education of the student. The rows identify the type of student being considered and the columns identify the type of friends, with panel 1 presenting the average and percentages for females and panel 2 for males. The bottom row shows the population shares of each group. The table is consistent with substantial homophily in friendship patterns over maternal education through the combined effect of sorting into schools and sorting into friendship. Looking along the diagonal of each panel, the percent of friends with the same maternal education as the student always exceeds the fraction of students in the population of that type. Females appear to exhibit higher levels of homophily than males at lower levels of maternal education.

Table 2 shows the same patterns by race and ethnicity. Again, the table is consistent with even higher levels of homophily since the fraction of own race friends far exceeds the fraction of that race in the population. Black and Hispanic females exhibit higher levels of homophily than black and Hispanic males. In general, Tables 1 and 2 also indicate that females have more friends than males, and students with college graduate maternal education have more friends than others. In order to separate the effect of school level segregation and homophily within schools, we also present the deviation of friendship frequencies within individual schools from expected friendship frequencies based on school level demographic composition. Appendix Table 1A and 2A confirm substantial homophily by maternal education and racial/ethnic groups within school.

Evidence Supporting the Research Design

To provide evidence that our use of across-cohort, within school variation is valid and uncontaminated by other unobservables, we conduct a series of balancing tests (following Bifulco et al. 2011, Lavy and Schlosser 2011, Billings et al. 2012) that estimate the associations between the cohort measures and individual-level exogenous attributes, such as age, health status, nativity status, etc. In Table 3, we regress cohort composition over maternal education, race and ethnicity on ten exogenous attributes of students, omitting the student themselves from this composition.[13] The specific cohort variables we use in both these balancing tests and our

[12] We also examine link models based on assuming a friendship exists when there is a link between the pair in at least one direction. The resulting estimates on the effect of demographic composition on link formation are very similar to the result presented here.

[13] These balancing tests follow Billings, Deming and Rockoff (2012) by reversing the regression relationship, as compared to Bifulco et al. 2011 and Lavy and Schlosser 2011, and placing the cohort composition on the left hand side so that a single F-test can be used to examine whether the set of exogenous attributes can systematically explain the within school by type variation associated with each cohort composition variable. Following Guryan et al. (2009) the balancing test models also control for school level composition omitting the student’s contribution in order to address the mechanical negative correlation between student's own attributes and cohort composition variables that omit the student. However, our cohorts are sufficiently large that the balancing tests results are very similar whether or not the Guryan et al. control is included in the models.

friendship formation model are percentage of black students, Hispanic students, students from "other" racial groups, students whose mom graduated from college, and students whose mom dropped out from high school, by grade-gender within schools. Each column in Table 3 represents a single regression of relating cohort composition on variables of individual characteristics of interest along with controls with school-gender fixed effects and cohort fixed effects.[14] Our results are consistent with cohort characteristics of interest that are conditionally plausibly exogenous (within schools) in that they cannot be explained by the predetermined attributes of the students in the cohort. Of the 60 individual t-tests, none is significant, and for the five regressions, none of the F-tests on the set of 10 variables is significant.

Estimating the Matching Model

Next, we describe the construction of our matched sample. For each student, we form a pair between him/her and each of the rest of the students from the same grade and gender. This process results in a fully matched sample of potential links in every school-cohort-gender cell. The size of the matched sample is about 12 million directed links, or 6 million unique pairs. For our friendship formation model, the outcome is a binary variable indicating whether the two parties in a pair nominated each other as their friend.

We defined four racial and ethnic categories (non-Hispanic white, non-Hispanic black, Hispanic, and Asians/Other race[15]) and 4 maternal education categories (four year college degree, high school graduate/some college, high school drop-out, and maternal education not reported). This implies 10 unique racial/ethnic combinations and 10 unique maternal education combinations of the two parties in a pair. Further, race/ethnicity and maternal education together define 16 student types. This results in 136 potential student-pair combinations or fixed effects for each gender g and school s.[16] Finally, in order to obtain a parsimonious vector β_t we restrict the interactions of pair type with cohort demographic composition so that cohort maternal education composition only affects friendship formation through the maternal education attributes of the pair of students, and, similarly, cohort racial and ethnic composition is restricted to only operate through the racial and ethnic attributes of the pair. That is, we do not allow interactions between the race types of the pair and cohort measures of maternal education levels.

Specifically, we estimate the effects of cohort composition on the likelihood of "types" of friendship pairs forming:

$$P_{ijtgcs} = \beta_{race}\left(\frac{Z_{gcs}^{race}}{n_{gcs}}\right) + \beta_{mom}\left(\frac{Z_{gcs}^{mom}}{n_{gcs}}\right) + \tilde{\delta}_{tgs} + \tilde{\mu}_{ijtgcs} \qquad (5)$$

where P_{ijtgcs} is the probability of a two way link between ego i with alter j and is a function of a large set of indicators variables reflecting the school and the potential pairs' type and interactions

[14] It is important to point out that the chance of a student nominating friends out of his/her own grade is not correlated with any of the cohort variables, suggesting that cross-grade friendship is not impacted by cohort composition.

[15] The majority of this group are Asian (70.46% indicate themselves not White, Black, Hispanic, Native American or other (not Asian). 56.06% clearly identify themselves as Asian), and results are robust to omitting non-Asians from this group. In some context below, we refer the "other" racial group as Asian when "other" may cause confusion.

[16] N(N+1)/2=4(4+1)/2=10; N(N+1)/2=(16*17)/2=136. An example of a pair type is white-dropout/white-college, indicating that one party of the pair is white with a high school dropout mom, and the other party of the pair is white with a college graduate mom.

between type and cohort-school composition in the type (for example, a pair type for race such as a white-African American pair is interacted with the proportion of black or Hispanic cohortmates). Z_{gcs}^{race} is the percentage of black, Hispanic and Other/Asian in a school-cohort-gender group (with g for gender, c for cohort/grade and s for school); Z_{gcs}^{mom} is the percentage of college graduate maternal education and high school dropout maternal education; n_{gcs} is the number of students in a school-cohort-gender cell. The large set of fixed effects constrain comparisons between individuals of the same pair type who attend the same school but are in different grade levels (cohorts) and are thus exposed to different cohort compositions. Because of the clear difference between male and female in Tables 1 and 2, we interacted all pair type indicators with a gender dummy and present the coefficients separately.[17]

Empirical Results—Matching Model

Table 4 presents the results related to the maternal education status of grade-mates. The estimate in each cell is the standardized coefficient from the interaction of a certain pair type dummy and cohort variables of maternal education.[18] The main finding is that, for females, increase in grademates with college educated mothers increases the likelihood of college grad/college grad pairs being friends as well as the likelihood of college grad/high school grad pairs. The F-statistic for the 40 parameters, which also include friendship links for individuals where maternal education is missing, is highly significant. For a sense of magnitude, the last two columns present the mean number of pairs of this type per cohort by gender and the implied number of students of each type associated with those pairs, e.g. 1,055 college-college pairs imply approximately 66 maternal college students.[19] A one standard deviation increase in the share of maternal college students scaled by cohort size is associated with a 0.3 percent point increase in the likelihood of friendship formation, or 3 additional friendships among those 66 students. The match sample average likelihood of friendship for a college-college pair is 1.3 percent so a one standard deviation change represents a 23 percent increase. Similarly, the estimate on maternal college-maternal high school pairs is associated with 9 additional friendships between 66 maternal college and 100 maternal high school students and a 36 percent increase over the base likelihood friendship of 0.9 percent.[20]

These results cannot be driven simply by an increase in the opportunity for college educated friends because we are estimating the probability of a specific link being formed. Rather, these results are consistent with an increase in the attractiveness of maternal college educated friends as the number of maternal college educated students rises, possibly because, given homophily, more individuals with college educated mothers leads to individuals with college educated mothers being more socially connected and therefore generating a greater social return associated with such friendships (Ballester et al. 2006). It is important to note that similar magnitude and same sign results exist for the college-college and college-high school pair variables interacted with percent high school drop-out mothers, but those are much less precisely

[17] As a check of our model, in Appendix Table 3A we show that our predicted number of mutual friends in total and by demographic categories are very close to the actual numbers at mean level.

[18] We present the unstandardized coefficients and standard deviation of scaled cohort composition variables in Appendix Table 4A and 5A.

[19] As a rough back of the envelope calculation, 1,035 non-directional pairs when omitting pairs with self implies 517 female pairs, and 33 female students implies 528=33*32/2 pairs and so is the integer that gets the closest to the 517 mean number of college-college pairs for a single gender.

[20] In Appendix Table 6A, we show the average likelihood of ties for each broad type of pair in the full sample of 12 million pairs.

estimated, potentially due to the small size of this group in the population. As a result, our key findings should be interpreted as the effect of an increase in the number of students with college graduate mothers in a cohort.

In Table 5 we present estimates of our matching model for the likelihood of various "types" of friendship links based on same-race or different-race matches. We find that increases in the share of blacks, Hispanics and Asians at the grade-level appear to increase homophily and decrease heterophily in friendship formation along specific dimensions. We find for females that increases in the proportion of black students in the grade increase same-race friendships for white and Other students (increases homophily) and reduces different-race friendships in white/Other and Hispanic/Other potential pairs (reduces heterophily). Increases in percent other/Asian lead to higher rates of black-black friendships. Again, the effects are non-trivial in magnitude. A one standard deviation increase in percent black leads to 25 additional friendships between whites and 7 additional friendships between female Asian students in our representative cohort with approximately 140 white and 36 Asian students.

Similarly, for males we find that an increase in the proportion of black students in the grade also increases the likelihood of same race links, for white and Hispanic pairs. Increases in the proportion of Hispanic grade-mates imply reductions in the likelihood of different-race pairs for Black/Other and White/Hispanic for females (reduced heterophily) and an increase in the likelihood of same-race pairs for black males (increased homophily). Increase in percent of students from other races also increases the likelihood of Hispanic-Hispanic friendships among males. Only the increase in the likelihood of black-white friendships as the proportion of 'other' race students increases operates in the opposite direction.

The main results from our matching model suggest that exogenous changes in the composition of class/schoolmates leads to changes in the likelihoods of the "types" of friend-pairs found in the data. More specifically, the results present direct evidence of increases in preferences for homophily relative to heterophily in this sample, especially with regards to race and ethnicity. Out of 10 statistically significant estimates on race and ethnicity, only one is associated with an increase across race friendships as share of a minority groups increases, yet it does not involve the racial group that increases (increase in the likelihood of black-white friendships with increases in percent of other race for males). While the literature has consistently found evidence of homophily, we know of no other work that documents this shift towards homophily as the population of minority groups increases. Further, in our analysis, the shift is identified using a quasi-random research design to estimate effects so that these changes cannot be attributed to other school level environmental changes that might often accompany equilibrium changes in demographic composition.

In order to examine the heterogeneity of friendship formation by school characteristics, we further split the schools in our sample into two even subgroups of students by high and low share of minority students, high and low share of students with maternal college education, and large and small schools. We then estimate the matching model within subsamples of pairs derived from the subgroups of schools respectively. The main findings from our matching model are quite similar in the stratified estimates.[21]

[21] In Appendix Table 7A-12A, we show the coefficients from stratified regressions. We also calculated the t-statistics for testing whether the relevant coefficients from two subsamples are significantly different from each other. We do occasionally observe signs of heterogeneity over school characteristics, but there is lack of clear pattern. In later sections, we also generate our instruments using stratified estimates from the match model to

Estimating the Effect of Friendship Composition on Academic Outcomes

Model Outline

Using our estimated model of friendship formation, we next develop predictions of friendship composition for individuals of any specific type in a specific cohort and school. For an individual *i* of type *x*, the predicted friendship outcome in terms of number of friends can be expressed based on summing the expression in equation (4) over all matches within the cohort by the individual type *y* of the potential friend j.

$$p_{ics} = \sum_{j \neq i, j \in \{c,s\}} (\beta_{xy} \left(\frac{Z_{cs}}{n_{cs}}\right) + \tilde{\delta}_{xys} + \tilde{\mu}_{ijxycs}) \quad (6)$$

By dropping the term involving the unobservables, we define the deterministic component of friendship outcomes as

$$\bar{p}_{xcs} = \sum_{j \neq i, j \in \{c,s\}} \left(\beta_{xy} \left(\frac{Z_{cs}}{n_{cs}}\right) + \tilde{\delta}_{xys}\right) \quad (7)$$

for any *i* of type *x*, since the deterministic component does not vary across individuals of the same type, school and cohort.

Similarly, using our model parameter estimates, we define the predicted friendship outcomes as

$$\hat{p}_{xcs} = \sum_{j \neq i, j \in \{c,s\}} \left(\hat{\delta}_{xys} + \hat{\beta}_{xy} \left(\frac{Z_{cs}}{n_{cs}}\right)\right) \quad (8)$$

where $\hat{\delta}_{xys}$ and $\hat{\beta}_{xy}$ are based on the model in equation (4). Again, the predictions do not vary across individuals of the same type, school, cohort and gender. The predicted number of friends of a given race, ethnicity or maternal education can be found by summing equation (8) over all matches within the cohort with students in that demographic category.

A key problem that arises from the estimation of $\hat{p}_{xcs}$ is that the estimates only vary across cohorts, $\hat{p}_{xcs} \neq \hat{p}_{xds}$ where $c \neq d$, if the total number of students of type *x* in school *s* is not large; otherwise cohort composition will simply represent school composition ($Z_{cs} \approx Z_{ds}$ for all cohorts *c* and *d* in a school). In fact, the estimates of $\hat{\beta}_{xy}$ on which $\hat{p}_{xcs}$ are based is only identified because Z_{cs} varies across cohorts. Therefore, while the total number of students of type *x* in the sample and the total number of students in any school or cohort may be relatively large, the number of students in each type in each school must be relatively small in order to create variation across cohorts. While our estimates of $\hat{\beta}_{xy}$ are consistent in the number of schools under the assumption of a linear probability model and the assumptions in equation (2), the dimensionality of our fixed effect vector increases linearly with the number of schools and the number of pair types, and so the school by student pair-type fixed effects suffer from an incidental parameters bias due to small numbers of observations in each cell. Specifically, the

---

examine whether our results for GPA are robust to the potential heterogeneity of friendship formation pattern (See Appendix Table 16A and 17A).

unobservable of a student *i* of type *x* in school *s* affects the estimates of $\hat{\delta}_{xys}$ for all types *y* and so the conditional expectation of the unobservable in the friendship choice equation $\tilde{\varepsilon}_{is}$ is non-zero.

$$E\left[\tilde{\varepsilon}_{is}\middle|\hat{\beta}_{xy}\left(\frac{Z_{cs}}{n_{cs}}\right), \hat{\delta}_{xys}\ \forall\ y\right] \neq 0 \tag{9}$$

In order to address this source of bias, we develop an individual specific measure of predicted friendship outcomes that explicitly omits all pairs involving individual *i* from fixed effects associated with pairs involving individuals of type *x*. This correction is based on the logic of using jackknife instrumental variable estimators for small samples (Angrist, Imbens and Krueger 1999; Blomquist and Dahlberg 1999) that develop a prediction for each observation using the entire sample, except the observation itself.

First, in a linear probability model, consistent estimates of $\hat{\beta}_{xy}$ can be and were obtained above by simply differencing out the school by pair type fixed effects in equation (5) and estimating

$$(P_{ijcs} - \bar{P}_{xys}) = \beta_{xy}\left(\frac{Z_{cs}}{n_{cs}} - \left(\overline{\frac{Z_{cs}}{n_{cs}}}\right)_s\right) + (\widetilde{\omega}_{ijcs} - \overline{\omega}_{cs}) \tag{10}$$

where $\widetilde{\omega}_{ijcs}$ is the sum of all terms involving unobservables in equation (5) and the bar operator implies the mean of the preceding term over all observations in a school by pair type cell.

In mean differenced models, the standard approach to estimating the fixed effects is to back out those fixed effect by calculating the mean of the within cell residual in the non-differenced sample. The individual specific fixed effect that omit pairs involving the individual *i* in cohort *c* can be estimated in the same way by summing the predicted residual over all cohorts *d* and pairs of students, *k* and *j*, with at least one student of type *x* other than student *i*.

$$\hat{\delta}_{xys}^{-i} = \sum\nolimits_{d\in\{s\}}\left(\sum\nolimits_{k\neq i\ if\ d=c, k\in\{x,d,s\}}\sum\nolimits_{j\neq k\ \&\ j\neq i\ if\ d=c, j\in\{d,s\}}\left(P_{xycs} - \hat{\beta}_{xy}\left(\frac{Z_{cs}}{n_{cs}}\right)\right)\right) \tag{11}$$

The notation in equation (11) is structured so that the first summation term sums over all cohorts in the school in order to calculate a school level fixed effect, the second term sums over all other students in the same cohort and of same type *x* as student *i*, and then the third term sums over all students of type *y* in the same cohort excepting students *i* and *k* if types *x* and *y* are the same.

Now based on equation (8), we define the individual specific prediction as

$$\hat{p}_{ics}^{-i} = \sum\nolimits_{j\neq i, j\in\{c,s\}}\left(\hat{\delta}_{xys}^{-i} + \hat{\beta}_{xy}\left(\frac{Z_{cs}}{n_{cs}}\right)\right) \tag{12}$$

However, as noted by Guryan, Kroft and Notowidigdo (2009), this process creates a negative correlation within type-cohort-school because an individual's contribution to the fixed effect is eliminated for themselves and not for anyone else in the type-cohort-school.[22] Guryan et

[22] In Guryan et al.' example, players select into golf tournaments, but are then randomly assigned to teams, which Guryan refers to as urns. The average team ability experienced by an individual golfer (omitting self) is negatively

al. proposes a solution to this bias for peer composition or subgroup means, which is to include an additional control for peer composition at a higher level of aggregation also omitting self. This control captures the negative correlation arising from omitting self and the estimates on the subgroup means are unbiased.[23]

In our context, students of a given type *x* sort into schools, but their allocation to a cohort or grade is assumed to be quasi-random. Therefore, the aggregate groups (or tournaments) are defined as type-school cells, and type-cohort-school cells are equivalent to one of Guryan et al.'s subgroups (or urns). We wish to separate the predicted friendship outcome from equation (8) into a component that omits all information involving choices made by individual i and a second component that contains this contamination.

$$\hat{p}_{ics} = \hat{p}_{ics}^{-i} + \hat{q}_{ics} \tag{13}$$

The expression $\hat{p}_{ics}^{-i}$ has been constructed in equation (12) so that it does not contain any information on the unobservable of individual i, and differencing equations (8) and (12) yields $\hat{q}_{ics}$

$$\hat{q}_{ics} = -\tau_{cs} \sum_{j \neq i, j \in \{c,s\}} (\hat{\delta}_{xys} - \hat{\delta}_{xys}^{-i}) \tag{14}$$

For our context, this contaminated component is equivalent to the control developed by Guryan et al. because it contains the contributions of the individual's choices to the conditional mean that is represented by the fixed effect estimates. The inclusion of this control will eliminate the bias caused by omitted an individual's own contribution to the fixed effect estimates in constructing predicted numbers of friends.

---

correlated (conditional on tournament fixed effects) with the individual's unobservable because within the tournament and urn the golfer cannot be paired with him/herself.

[23] In order to apply the Guryan et al. logic to our example, it is useful to consider a slight generalization to their problem. Consider the following simple behavioral model

$y_{ics} = \beta X_{cs} + \delta_s + \pi_{ics}$

where *c* is an urn and *s* is a tournament. Assume that for any individual *i*, X_{cs} is correlated with ε_{ics} , but can be divided into two additively separable components

$X_{cs} = X_{ics}^{i} + X_{ics}^{-i}$

where the first component contains the contamination that leads to the correlation and the second component is uncorrelated with ε_{ics}

$E[\pi_{ics} | X_{ics}^{i}, \delta_s] = \alpha X_{ics}^{i}$

$E[\pi_{ics} | X_{ics}^{-i}, \delta_s] = 0$

The second component X_{ics}^{-i} is equivalent to the average urn ability omitting self, and simply including this control will lead to biased estimates because X_{ics}^{i} is omitted and X_{ics}^{i} and X_{ics}^{-i} are correlated. However, as suggested by Guryan et al., including both variables yields unbiased estimates since

$E[y_{ics} | X_{ics}^{-i}, X_{ics}^{i}, \delta_s] = \beta X_{ics}^{-i} + (\beta + \alpha) X_{ics}^{i} + \delta_s$

While the Guryan et al. idea of controlling for the tournament mean minus the individual's contribution seems intuitively appealing, the true source of the solution is that the within tournament variation in this mean nearly perfectly correlates with the individual, additively separable portion of the mean (the contaminated component) that has been removed from the variable of interest.

Finally, consider an empirical model of an outcome y_{ics} where a student of type *x*'s outcome may be influenced by the type of social links formed by the student:

$$y_{ics} = \theta \underline{p}_{ics} + \gamma_{xs} + \tau_{ics} \tag{15}$$

where $\underline{p}_{ics}$ is a vector of friendship composition outcomes, such as number of friends and number of friends of different demographic groups, γ_{xs} is a vector of school by student type fixed effects, and p_{ics} potentially correlates with the unobservable τ_{ics}.

Therefore, we estimate a series of first stage models where the friendship composition outcome depends upon the individual level prediction of composition, a second term containing the contaminated component of the prediction, and the school by type fixed effects.

$$p_{ics} = \omega_1 \hat{p}_{ics}^{-i} + \omega_2 \hat{q}_{ics}^{-i} + \varphi_{xs} + \tilde{\rho}_{ics} \tag{16}$$

where any element of p_{ics} depends only on the same elements of $\hat{p}_{ics}^{-i}$ and $\hat{q}_{ics}^{-i}$, e.g. number of friends or number of friends whose mothers have a college education, so that the coefficients in equation (17) are scalers.

We propose to obtain consistent estimates of θ in equation (15) using a second stage estimation equation based on the estimates of equation (16) as follows

$$y_{ics} = \theta \underline{\hat{p}}_{ics} + \pi \underline{\hat{q}}_{ics}^{-i} + \tilde{\gamma}_{xs} + \tilde{\tau}_{ics} \tag{17}$$

where this equation also includes the predicted composition based on equation (16) and the contaminated component of the instrument in order to avoid the Guryan et al. bias.

Data Description

In the following two sections, we first present the descriptive statistics of the student level data relevant to our examination of friendship effects of GPA. Then we present our estimates. After describing the data, we begin the empirical analysis by presenting standard OLS models that links the GPA of friends together. However, these models are likely biased due to the endogeneity of friends. We next show that, using our matching model from the previous section, we can predict the "types" of friends that individuals nominate in the data using across-cohort variation in the "supply of friend-types". We then incorporate our predicted friendship patterns as instruments in a two-stage analysis to examine the importance of endogeneity. As we show above in the context of our matching model, we also present balancing test results that show that individual covariates are unrelated to our instruments.

Table 6 shows the means of GPA by maternal education and racial/ethnic groups at the student level. Average GPA is lower among students with lower maternal education. The mean level of GPA among students with missing maternal education is close to the GPA in the group of high school dropout. For all racial/ethnic groups, female students have higher average GPA than males; black and Hispanic students show lower GPA than the other two groups. We also provide pooled descriptive statistics for the key variables used in our analyses in Appendix Table 13A. Among students in our sample, 91% were born in the U.S., 92% report living with their mother and the average family size is 4.3 persons per household.

Empirical Results—Effects of Friends on Academic Achievement

Our next step is to leverage the predicted friendship pattern measures we extract from the matching model above to use as instruments for actual friendship patterns. Like any instrument, our measures need to be strongly related to the endogenous (actual) friendship pattern and unrelated to the unobservables determining GPA. In Table 7 and Appendix table 14A, we show that our predicted friendship composition measures are strongly related to the actual friendship nominations in the data, where the F-statistics are between 30-240, even after controlling for school by type and school by cohort fixed effects and eliminating any effect of individual's own friendship choices. In Table 8, we show that our instruments are unrelated to a large set of observable factors ("balancing tests," similar as in Table 3), which is consistent with the exclusion restriction.[24]

Table 9 presents estimates of the effects of friend composition of maternal education on students' GPA for female and male sub samples. Each column represents results from a single regression with school-type FE and school-grade FE. The OLS coefficients from column (1) and (4) shows students having more friends with a college educated mother have higher GPA relative to their grade mates, but having more friends whose mom dropped out from high school doesn't significantly correlate with lower GPA. The pattern shows no gender difference.

In Columns 2, 3, 5, 6 in Table 9, we then examine friend composition effects for academic achievement using two-stage least squares. Considering the low number of mutual friends on average, we test one "type" of friend at a time. For example, in column (2), we regress GPA on the actual number of friends with a college graduate mom, instrumented by the predicted number of friends whose moms graduated from college.[25] The first observation is that the IV estimates differ from OLS estimates. The coefficient of college graduate mom increases by 39% for female, but changes from significant and positive to insignificant, small and negative for males. The coefficient of dropout from high school remains relatively small and statistically insignificant. As noted above, the F-stat from first stage of the 2SLS is in the range of 30-240, by which we can reject the null hypothesis of weak instruments.

The estimated coefficient of peer college graduate mom is 0.212, indicating that one more mutual friend with a college educated mom is associated with a 0.21 grade point increase of GPA, which is about a 7.3 percent increase at a mean GPA of 2.88 for all female students and represents a 0.272 standard deviation increase in GPA. Multiplying by the standard deviation of number of maternal college friends, the effect of a one standard deviation increase in the number of friends of this type is a 0.178 standard deviation increase in GPA. In contrast, the number of friends with mom dropped out from high school is not significantly correlated with GPA. Further, for males, maternal education of friends does not show any significant effect on own

[24] As above, the balancing test regressions include the Guryan control to avoid the downward bias caused by omitting self from the construction of the instruments. The balancing test results are also robust to reversing the regressions so that the attribute is regressed on the instrument and the control as in Guryan, Kroft and Notowidigdo (2009).

[25] Relevant Guryan type controls are always included in both first and second stages.

GPA in the IV regression, in contrast to the OLS regression. The small and insignificant IV coefficient for males must be interpreted with some caution because the IV standard errors for the effect of having friends whose mothers are college graduates is substantially larger for the male than the female sample.

As we discuss above, a remaining question about our estimated effect of friend characteristics on own GPA is whether the cohort composition of maternal education directly operates on girls' GPA, instead of only through the effect of having friends with college educated mothers. In order to partially address this concern, we examine whether this "direct environmental effect" can be found in the male sample[26]—that is, if a concern is that cohort composition has more general, direct effects on own GPA for females, one way to indirectly test this effect is the cohort composition measures for females also predict GPAs for the males in the same school/cohort. We conduct this examination by using the parameters from the friendship match model for the female sample to predict the number of friends (by type) for males, and then use these predicted parameters for males to predict friends. We then use instruments generated from the experiment to re-estimate our GPA model. We find the female's match model does not predict male's GPA or improve the precise of the estimates from the male IV model for GPA. The instrument generated from the male's match model also does not predict female's GPA, with the coefficient for maternal college friends becoming negative and insignificant. All coefficients from this inverse-gender experiment are not statistically different from zero.

To further examine the robustness of our main results, we examine the sensitivity of the results to the specification of the IV model, and whether other aspects of friendship composition might directly influence academic outcomes and simply be correlated with maternal education of friends. In Appendix Table 15A, we present results from a set of IV models. First, instead of testing one instrument at a time as in previous tables, we explore whether including both friends with high maternal education and low maternal education influences our results. Then we also examine the effect of controlling for the total number of friends and the racial and ethnic composition of friends.[27] The positive effect of friends with high maternal education on own GPA for female is robust through the three specifications we test (column 1-3). Still, no distinguishable impact of friends' maternal education is found for males. Adding the total number of friends does not change the pattern of correlation between friends composition and own GPA for either females or males. None of the coefficients of number of black or Hispanic friends is significantly different from zero, suggesting that the effect from racial composition of friends is quite weak when controlling for friends' maternal education.

We also generate our instruments using stratified estimates from the match model to examine whether the potential heterogeneity of friendship formation pattern would affect our findings. Our results for GPA are robust for all three ways of splitting the pair sample by school characteristics, as shown in Appendix Table 16A, and so our estimated effects for girls do not appear sensitive to the particular friendship link model estimated. We then focus on the female GPA and examine the heterogeneity of our IV results by student level subsamples by school

[26] We thank Damon Clark for suggesting this direction of examining the exclusion restriction in our main models.

[27] The first stage results for additional friendship variables are shown in Table 15A. Note that univariate IV models are used for simplicity and ease of interpretation, but our two stage results are robust to an IV model where all instruments are used to create predicted values for all variables.

characteristics in Appendix Table 17A. The effect of maternal college friends on GPA is concentrated in schools with low share of minority students, schools with low share of college educated mothers, and small schools.

Mechanisms Analysis

Next, in order to further examine the overall GPA effects for females, Table 10 decomposes the result based on the four subject areas of grades available in the data (Math, English, Science, and History). The evidence suggests that the gain in GPA from having a friend with a highly educated mother is based on better performance in both English and Science classes, but not Math and History Courses. Mirroring the main results, we find no effects for males. We also examined whether the number of friends by maternal education predicts students' choice of taking certain subject of courses, and found no relation.

In order to investigate the potential channel through which girls are affected by close friends with high maternal education, we use the preferred IV specification to examine a series of non-cognitive outcomes. Given our interest in identifying consistent patterns of results, we also indicate findings that are significant at the 10 percent level for this analysis. To reduce the number of tests, we manually classify variables into seven categories and then use factor analysis procedures to generate composite variables for those categories—self evaluation,[28] judgment regarding social environment,[29] mental status, trouble in school activities, misbehavior, smoking and drinking, and self reported health. A high score reflects high self evaluation, comfortable social environment, good mental health, having more trouble at school, more misbehavior, high frequency of smoking/drinking and good physical health respectively. More details of the variables included in the factor analyses are in Appendix Table 18A.

In Table 11, we present the results of our mechanisms analyses. Each column of Table 11 refers to a single outcome of interest, and each cell represents the relevant coefficient of interest from a separate IV regression. The results suggest that female's subjective evaluation regarding self and school are consistently positively correlated with the number of friends with high maternal education they have. Female students with more friends of high maternal education are more confident and comfortable with themselves and the people around them. The results also indicate that girls with more friends whose moms graduate from college are less likely to display depression symptoms or misbehave/act out in school. The findings for self-evaluation and social environment are most notable because there is little or no relationship between these variables and friends' maternal education for male students. On the other hand, the smoking/drinking index is associated with maternal college for male students, while these health behavior oriented variables have little or no relationship maternal college graduate for the female sample.

Our results support that girls are more influenced by high quality peers than boys on self evaluation and social comfort, as suggested in relevant previous literature (Brown 1982, Griffin

[28] Self-evaluation covers rating to questions including whether the interviewees think themselves physically fit, are proud of themselves, like themselves, think they are doing things right, and try to study well. When discussed in Psychology, the concept of self-esteem often needs to be clarified—either a general term on overall feeling about self or on a specific aspect, such as academic related, physical appearance and social popularity, etc. It is also an important indicator for troublesome behavior and depression (Rosenberg et al. 1989, Markowitz 2001).

[29] Environment evaluation shows the extent that students feel close, safe, fair and accepted at school.

et al., 1999, “role model effect” in Durlauf (2004)), but less likely to be influenced in terms of exhibiting problematic behaviors. We also run correlation analysis and confirm that low self evaluation, passive attitude and behavior at school and poor mental health are negatively associated with GPA in our sample, even after removing school by cohort and school by student type fixed effect. Numerous studies in education have found that academic achievement and self-esteem are positively correlated (Bankston and Zhou, 2002; Ross and Broh, 2000; Schmidt and Padilla, 2003; Wong and Watkins, 2001). Purky (1970) argued that there is continuous interaction between self-esteem and academic achievement. Byrne (1984) reviewed the empirical findings in this literature, both cross-sectional and longitudinal designs, and also confirmed the existence of the relationship. Our analysis is novel because we have plausibly exogenous variation in friendship composition that can separate correlational and causal effects. However, the causal link between self-esteem and school achievement is still under debate. Some investigators argue that high education achievement and self control enhance self esteem, not vice versa, and our analysis cannot shed light on this debate because, as with the mechanism analyses in earlier cohort studies (Bifulco et al. 2011, Lavy and Schlosser 2011), we have only shown that friendship composition has a causal influence on both the variable of interest and on the potential mechanism and not whether one of these effects operates through the other.

As a further test of the relevance of these potential mechanism variables, we re-estimate our two-stage IV model for girls allowing the effect of predicted friendships to vary across the three maternal education subgroups: maternal college educated, maternal high school graduate and maternal high school drop-out. The resulting estimates for GPA are shown in the first column of Table 12 and imply that most of the effect of maternal college friends on GPA is concentrated in the maternal college educated subsample with an effect of 0.281 approximately 33 percent larger than the estimate for the full sample. We observe a statistically insignificant positive effect of 0.112 and no effect for the maternal high school drop-out subsample. We then re-estimate the models for the mechanism variables finding that the positive effects of maternal college on most of these variables (social comfort, mental health and misbehave) is also concentrated among the maternal college sample, and to a much lesser extent in the maternal high school graduate sample, with fewer effects in the maternal drop-out sample.

The one exception is the self-evaluate index that we associate with self-esteem. Self-esteem is improved by maternal college friendships for all groups with the largest effects for students who mothers did not graduate from high school, a group that did not show any improvement in GPA with maternal college friends. However, increases in self-esteem may be important on its own especially for girls whose mothers did not graduate from high school, and these effects are sizable with a one standard deviation increase in number of maternal college friends leading to a 0.283 standard deviation in self-esteem for students with drop-out mothers.[30]

Calibration of the Effects of Friendship Patterns on GPA

Next we conduct a simple calculation of the effect of our key significant findings on the effect of educational composition on friendship formation. Specifically, in Table 4, we show that an increase in the share of maternal college educated students increases the likelihood of a

[30] The coefficient is 0.443 from Table 12. The sd. of self-evaluate is 1.026 for female sample. The sd. of number of maternal college friends for female sample is 0.655.

match between two students who both have a college educated mother as well as the likelihood of a match between two students where one has a college educated mother and the other has a mother who is a high school graduate. In our sample, we increase the share of maternal college educated students in every cohort by 10 percentage points and then examine the direct effect of this increase on the predicted number of college educated for students overall and by level of maternal education.

The predicted number of maternal college educated friends changes with number or share of maternal college students for two reasons. First, there are simply more potential friendship matches available with students whose mothers are college graduates. Second, the probability of matches or links increases both between two maternal college students and between a maternal college and a maternal high school graduate student based on the statistically significant estimates on percent maternal college in Table 4 and 4A.

Assuming that cohort size is held constant, the number of maternal college educated students after the change (N_{2c}) is simply

$$N_{2c} = N_{1c} + 0.1N \tag{18}$$

where N_{1c} is the initial number of maternal college educated students, and N is the total number of students in the cohort. The resulting change in number of maternal college friendship links or matches for a maternal college educated student (D_c) is then

$$D_c = (N_{1c} + 0.1N - 1)\left(P_{cc} + \beta_{cc}\frac{0.1}{N}\right) - (N_{1c} - 1)P_{cc} \tag{19}$$

where P_{cc} is the probability of a link and β_{cc} is the estimated coefficient on the share maternal college educated for college-college links. This expression can be rewritten to illustrate the separate effects of changes in the probability of a link and changes in the number of potential links

$$D_c = (N_{1c} + 0.1N - 1)\left(\beta_{cc}\frac{0.1}{N}\right) + (0.1NP_{cc}) \tag{20}$$

For maternal high school graduate students the change in maternal college friends (D_h) is

$$D_h = (N_{1c} + 0.1N)\left(\beta_{ch}\frac{0.1}{N}\right) + (0.1NP_{ch}) \tag{21}$$

where P_{ch} is the probability of a link and β_{ch} is the estimated coefficient on the share maternal college educated for college-high school links. Finally, for maternal high school drop-out students we set the parameter estimate on share maternal college to zero due to the small and statistically insignificant estimate and the predicted change is

$$D_d = 0.1NP_{cd} \tag{22}$$

where P_{cd} is the probability of a link.

In order to calculate these expressions for the sample, we use the within school sample average frequencies of link formation between potential links for P_{cc}, P_{ch} and P_{cd}. We set P_{cc}, P_{ch} and P_{cd} to the empirical frequencies observed in each school so that our calculations capture the fact that schools differ in the likelihood of link formation due to, for example, across school differences in the racial and ethnic composition of each maternal education subgroup. Further, the use of the empirical frequencies is consistent with holding cohort size constant because one would expect link frequencies to fall on average as cohort size increases. These results are shown in Column 1 where the rows present the results for the overall sample, the maternal college subsample, the maternal high school subsample and the maternal drop-out subsample. A 10 percentage point increase in the share of maternal college students increases the sample average fraction of students who have a mother with a college degree by 34 percent over an original base fraction of 0.292.[31] The average number of predicted maternal college friends increases by 0.238 from a base of 0.316 or by 75 percent. The maternal college subsample has an increase of 0.243 over a base of 0.543 or 45%, and the maternal high school graduate sample has an increase of 0.317 over a base of 0.299 or 106%. The maternal high school drop-out sample has the smallest absolute increase of 0.137, but the largest percent increase of 109% over a base of 0.126.

Notably, the percentage increases in maternal college friends are substantially smaller for maternal college educated students than for maternal high school graduate or drop-out students. While this seems surprising given the strong effects of percent students with maternal college graduates on the likelihood of link formation between two students with maternal college graduates, the increase in the likelihood of college-college and college-high school links explains only a moderate fraction of the increased average number of friends, 0.049 for the maternal college subsample and 0.042 for the maternal high school graduate subsample.[32] The primary driver of the increase in the number of maternal college friends for all groups is the increasing number of friendship opportunities. This effect is smallest for maternal college students because the percent increase in maternal college students is smallest in the cohorts that have the largest share of maternal college students. In the maternal college subgroup, observations are more likely to come from schools with a larger share of maternal college students than average and so the smallest percent increase in maternal college friendship opportunities. Further, given the strong negative correlation between the presence of maternal college students and maternal high school drop-out students, the largest percentage increases in maternal college friendship opportunities occur for the maternal high school drop-out subsample in the schools with the largest maternal high school drop-out population.

In terms of estimating the impact on GPA, the first stage effect of predicted number of friends on actual friends is 0.860, and then the effect on GPA is 0.212 grade point from the instrumental variable analysis. Therefore, our calculations suggest that the direct effect associated with a 10 percentage point increase in the share of maternal college students in each cohort increases girls GPA by 0.043.[33] Turning to the maternal education subsamples, we draw on the estimated effects in Table 12 where we observed large positive and significant results for

---

[31] The increase is 41% if the fraction of maternal college students is calculated based on all four maternal education categories: college, high school, drop-out and missing, rather than omitting missing from the calculation.

[32] Given the base numbers of friends above, the percent increase for maternal college students is 9 percent and the percent increase for maternal high school graduate students is 14 percent comparable to, but somewhat larger than, the back of the envelop calculations presented earlier of 7 and 10 percent. The large estimates arise in part because the effect of the increased probability of link formation is calculated for a larger number of maternal college students as shown in the first terms of equations (20) and (21).

[33] This result arises from multiplying the predicted change in number of maternal college educated friendships times the first stage instrumental variables coefficient in order to obtain the correct scale and then multiplying this product by the coefficient on GPA from the second stage model or 0.238*0.860*0.212=0.043. The subsample calculations follow the same form except use the change in number of maternal college educated friends for each subsample.

the maternal college subsample and insignificant, but appreciable, effects for the maternal high school subsample. For the calculated increases in number of maternal college friends, the increases in GPA are 0.059 for maternal college students and 0.031 for maternal high school students.[34] Finally, the large effect estimates for students whose mothers did not high school imply significant changes in self esteem for this group, with the self esteem rising by 0.033 standard deviations for maternal high school dropout, which is comparable to the 0.033 and 0.041 for maternal college students and maternal high school students, respectively.[35]

Conclusions

This paper presents new evidence of the determinants of friendship links and the effects of the characteristics of friends on own school achievement. We use a novel strategy that leverages across-cohort, within school variation in the "supply of friend 'types'" for both sets of results. We first show that small variations in the supply of friends increase homophily and reduce heterophily in friendship formation patterns in high school. This is consistent with both the biological evidence that individuals prefer to have friends like themselves as well as the large body of empirical work that shows strong correlations in the characteristics of friends (i.e. homophily). However, we are the first to examine these effects within a quasi-experimental research design[36] and to provide evidence of how the pattern of homophily increases as the population of minority groups increase. These results have strong implications for policies that attempt to "rewire" social networks by increasing the opportunities for choosing friends who are different. Our results suggest that increasing opportunities may not be enough to foster heterophilious friendships, which is also a likely explanation for the results from Carrell et al. (2013), where randomizing individuals into squadrons in the military to foster heterophily actually reduced the performance of the heterophilious group.

We then use our predictions of friendship formation to leverage a second research question—whether having friends with highly educated mothers is causally related to academic achievement or whether the correlation in GPA between friends is a result of endogenous friendship selection. We find both cases—for female high school students, our results suggest that increases in friend maternal education status leads to large GPA increases, which are concentrated in coursework in science and English. We also find that the OLS estimates for males are driven by endogeneity, and once corrected are small and no longer statistically significant.

In order to examine the mechanisms linking the maternal education of friends to own academic achievement, we show evidence that, for females but not males, friend maternal education is also linked to reductions in increases in feelings of self worth and favorable opinions of the school environment. Having more friends with college educated parents appears

---

[34] For maternal college students, the calculation is 0.243*0.860*0.281=0.059. For maternal high school graduate students, the calculation is 0.317*0.860*0.112=0.031.

[35] Since the absolute magnitude of mechanism factors does not have realistic meaning, we are only presenting the standardized effect for mechanism factors. For maternal high school dropout sample, the calculation is ((0.137*0.860*0.443)/1.026)*0.655=0.033. For maternal high school graduate sample, the calculation is ((0.317*0.860*0.234)/1.026)*0.655=0.041. For maternal college sample, the effect is ((0.243*0.860*0.250)/1.026)*0.655=0.033.

[36] The most similar work examines dating patterns rather than high school friendship formation (Fisman et al. 2008).

to lead to both higher levels of self-esteem and higher grades among girls while friendship composition does not appear to affect other significant intermediate outcomes like disciplinary problems or health outcomes. We also show that both our findings for GPA and our findings for the mechanism variables, with the exception of self-esteem, are concentrated among the subsample of students whose mothers have a college degree. While the subsample findings work against self-esteem as a mechanism, they point to an important effect of friends for low maternal education girls, namely maternal college educated friends create large increases in self-esteem for this subsample.

Finally, we conduct a series of calculations in order to illustrate the effect of small to moderate increases in share of students with maternal education at the college level. The key findings from this calibration are: first, that the increased opportunities for additional friendships with maternal college educated students dominates the direct effect associated with changes in the probability of friendship formation as the share of maternal college students increases; second, the opportunity effects are largest among students with less educated mothers where the percent increase in share maternal college are the largest; and third, the students with high maternal education will benefit the most in GPA and in other aspects of well-being from the assumed increase in the share of maternal college educated students. The one exception to this conclusion is the effect of maternal college friendships on self-esteem for girls whose mothers did not graduate from high school.

References

Angrist, Joshua, Guido Imbens, and Alan B. Krueger. "Jackknife instrumental variables estimation." *Journal of Applied Econometrics*, 15 no.1 (1999): 57–67.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. "Who's who in networks. wanted: the key player." *Econometrica* 74, no. 5 (2006): 1403-1417.

Bankston, Carl L., and Min Zhou. "Being Well vs. Doing Well: Self-Esteem and School Performance among Immigrant and Nonimmigrant Racial and Ethnic Groups." *International Migration Review* 36, no. 2 (2002): 389-415.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross. "The effect of classmate characteristics on post-secondary outcomes: evidence from the Add Health." *American Economic Journal: Economic Policy* 3, no. 1 (2011): 25-53.

Billings, Stephen B., David J. Deming, and Jonah E. Rockoff. *School Segregation, Educational Attainment and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg*. No. w18487. National Bureau of Economic Research, 2012.

Blomquist, Sören, and Matz Dahlberg. "Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments." *Journal of Applied Econometrics* 14, no. 1 (1999): 69-88.

Brown, B. Bradford. "The extent and effects of peer pressure among high school students: A retrospective analysis." *Journal of Youth and Adolescence* 11, no. 2 (1982): 121-133.

Byrne, Barbara M. "The general/academic self-concept nomological network: A review of construct validation research." *Review of educational research* 54, no. 3 (1984): 427-456.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. "Peer effects and social networks in education." *The Review of Economic Studies* 76, no. 4 (2009): 1239-1267.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81, no. 3 (2013): 855-882.

Durlauf, Steven N. "Neighborhood effects." *Handbook of regional and urban economics* 4 (2004): 2173-2242.

Fisman, Raymond, Sheena S. Iyengar, Emir Kamenica, and Itamar Simonson. "Racial preferences in dating." *The Review of Economic Studies* 75, no. 1 (2008): 117-132.

Fletcher, Jason M., and Stephen L. Ross. *Estimating the Effects of Friendship Networks on Health Behaviors of Adolescents*. No. w18253. National Bureau of Economic Research, 2012.

Griffin, Kenneth W., Gilbert J. Botvin, Margaret M. Doyle, Tracy Diaz, and Jennifer A. Epstein. "A six-year follow-up study of determinants of heavy cigarette smoking among high-school seniors." *Journal of behavioral medicine*22, no. 3 (1999): 271-284.

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo. “Peer effects in the workplace: Evidence from random groupings in professional golf tournaments.” *American Economic Journal: Applied Economics* 1, no.4 (2009): 34-68.

Haveman, Robert, and Barbara Wolfe. "The determinants of children's attainments: A review of methods and findings." *Journal of economic literature* 33, no. 4 (1995): 1829-1878.

Hoxby, Caroline. *Peer effects in the classroom: Learning from gender and race variation*. No. w7867. National Bureau of Economic Research, 2000.

Lavy, Victor, and Edith Sand. *The Friends Factor: How Students’ Social Networks Affect Their Academic Achievement and Well-Being?*. No. w18430. National Bureau of Economic Research, 2012.

Lavy, Victor, and Analia Schlosser. "Mechanisms and impacts of gender peer effects at school." *American Economic Journal: Applied Economics* 3, no. 2 (2011): 1-33.

Manski, Charles F. "Identification of endogenous social effects: The reflection problem." *The review of economic studies* 60, no. 3 (1993): 531-542.

Manski, Charles F. *Identification problems in the social sciences*. Harvard University Press, 1995.

Manski, Charles F. “Economic analysis of social interactions.” *Journal of Economic Perspectives*14, no.3 (2000): 115-136.

Markowitz, Fred E. "Modeling processes in recovery from mental illness: relationships between symptoms, life satisfaction, and self-concept." *Journal of Health and Social Behavior* (2001): 64-79.

Mayer, Adalbert, and Steven L. Puller. "The old boy (and girl) network: Social network formation on university campuses." *Journal of public economics* 92, no. 1 (2008): 329-347.

Moody, James. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107, no. 3 (2001): 679-716.

Purkey, William Watson. "Self concept and school achievement." (1970).
Rosenberg, Morris, Carmi Schooler, and Carrie Schoenbach. "Self-esteem and adolescent problems: Modeling reciprocal effects." *American sociological review* (1989): 1004-1018.

Ross, Catherine E., and Beckett A. Broh. "The roles of self-esteem and the sense of personal control in the academic achievement process." *Sociology of Education* (2000): 270-284.

Schmidt, Jennifer A., and Brenda Padilla. "Self-esteem and family challenge: An investigation of their effects on achievement." *Journal of youth and adolescence* 32, no. 1 (2003): 37-46.

Weinberg, Bruce A. *Social interactions with endogenous associations*. No. w13038. National Bureau of Economic Research, 2007.

Wong, Martha Suk Wai, and David Watkins. "Self-esteem and ability grouping: A Hong Kong investigation of the big fish little pond effect." *Educational psychology* 21, no. 1 (2001): 79-87.

Table 1. Number and Distribution of Mutual Friends by Maternal Education

		Friend's Maternal Education			
	No. of Friends	High School Dropout	High School Graduate	College Graduate	Missing
Own Maternal Education					
Female					
High School Dropout	0.888	23.3%	45.8%	14.1%	16.8%
High School Graduate	1.171	9.3%	53.6%	25.8%	11.3%
College Graduate	1.302	4.9%	43.7%	42.0%	9.3%
Missing	0.795	13.7%	45.3%	22.1%	19.0%
Male					
High School Dropout	0.524	17.5%	45.5%	16.6%	20.4%
High School Graduate	0.732	6.9%	50.0%	28.8%	14.3%
College Graduate	0.866	3.4%	39.5%	45.8%	11.3%
Missing	0.469	9.4%	43.4%	24.7%	22.5%
Maternal Education	Obs.				
Distribution	84,695	10.7%	44.3%	25.4%	19.7%

Note: A mutual friend tie is defined as a two-way nomination of friendship. The "No. of friends" column presents mean of the number of mutual friends occurred to students in each category of maternal education by gender. The other four columns present the mean number of mutual friends by friends' maternal education as a percentage of the total number of mutual friends. The last row includes the total sample size and distribution of students by maternal education. In the following tables, we will label high school dropout as "HS dropout", high school graduate as "HS grad", and college graduate as "College Grad" when space is limited.

Table 2. Number and Distribution of Mutual Friends by Racial Groups

		Friend's Race			
	No. of Friends	White	Black	Hispanic	Other
Own Race					
Female					
White	1.297	87.7%	1.6%	5.3%	5.4%
Black	0.884	7.0%	80.6%	7.2%	5.1%
Hispanic	0.797	28.2%	8.8%	55.4%	7.6%
Other	0.960	41.1%	8.9%	10.7%	39.3%
Male					
White	0.840	86.0%	1.7%	5.7%	6.7%
Black	0.453	10.9%	73.8%	9.4%	5.8%
Hispanic	0.464	32.5%	8.4%	49.5%	9.6%
Other	0.632	43.2%	5.8%	11.1%	39.9%
	Obs.				
Race Distribution	84,695	55.2%	17.1%	17.1%	10.7%

Note: The “No. of friends” column presents mean of the number of mutual friends occurred to students in each racial/ethnic group by gender. The other four columns present the mean number of mutual friends by friends’ racial group as a percentage of the total number of mutual friends. The last row includes the total sample size and distribution of students by racial groups.

Table 3. Balancing Test for Cohort Composition Sorting with Student Characteristics

Independent Variables	%Black (1)	%Hispanic (2)	%Asian (3)	%Mom College Graduate (4)	%Mom HS Dropout (5)
Age	-0.00019	-0.00024	0.00000	-0.00009	0.00024
	(0.00041)	(0.00031)	(0.00022)	(0.00032)	(0.00026)
No. of People in Household	0.00021	0.00003	0.00013	-0.00001	0.00010
	(0.00015)	(0.00016)	(0.00010)	(0.00020)	(0.00014)
No. of School Kids in Household	0.00003	0.00011	-0.00023	0.00020	0.00020
	(0.00018)	(0.00018)	(0.00014)	(0.00020)	(0.00016)
Live with Both Parents	-0.00011	-0.00029	-0.00011	0.00049	-0.00058
	(0.00044)	(0.00035)	(0.00021)	(0.00049)	(0.00035)
Live with Biological Parents	-0.00067	-0.00013	0.00013	0.00078	-0.00002
	(0.00075)	(0.00063)	(0.00036)	(0.00095)	(0.00066)
Mother's Edu in Single Year	-0.00003	-0.00005	0.00009	0.00018	-0.00003
	(0.00012)	(0.00006)	(0.00005)	(0.00017)	(0.00010)
Mother Born in US	0.00181	-0.00049	-0.00082	-0.00003	-0.00015
	(0.00111)	(0.00047)	(0.00066)	(0.00066)	(0.00053)
Born in US	-0.00061	-0.00045	-0.00031	-0.00130	-0.00011
	(0.00065)	(0.00068)	(0.00054)	(0.00067)	(0.00069)
Adopted	-0.00125	0.00037	0.00012	0.00003	-0.00058
	(0.00108)	(0.00090)	(0.00078)	(0.00102)	(0.00088)
Health Condition at Birth	0.00016	0.00010	0.00118	0.00110	-0.00022
	(0.00099)	(0.00105)	(0.00078)	(0.00128)	(0.00097)
Observations	84,689	84,687	84,689	84,680	84,680
R-squared	0.976	0.976	0.93897	0.904	0.890
Ftest	0.852	0.639	1.523	1.077	0.797
Fpvalue	0.580	0.778	0.137	0.384	0.632

Note: Each column displays a separate regression of a cohort composition variable on ten predetermined demographics variables. To maintain the sample size, we also include interaction terms of each predetermined demographic variable and the indicator for non-missing value of that variable. The cohort composition variables for a student includes the percentage of black (not Hispanic), Hispanic, other/Asian, mother graduated from four year college and mother dropout from high school, omitting the student's own contribution. All regressions control for school-gender fixed effect, grade dummies, and Guryan type control for school level composition omitting the student him/herself. Standard errors are clustered at the school level. Observations with missing maternal education data are assigned the median value of the cohort variable of all other students in the school-grade-gender group. **p<0.01 and *p<0.05.

Table 4. Friendship Pattern Standardized Estimation from Matching Model for Maternal Education

Maternal Education Pair Type	Female % College Grad (1)	Female % HS Dropout (2)	Male % College Grad (3)	Male % HS Dropout (4)	Mean No. of Pairs in Cohort by Gender	Implied No. of Students
College Grad-College Grad	0.00304**	0.00388	0.00084	0.00063	1055	33
	(0.00077)	(0.00443)	(0.00067)	(0.00219)		
College Grad-HS Grad	0.00328**	0.00268	0.00105	0.00111	2774	
	(0.00055)	(0.00222)	(0.00076)	(0.00096)		
College Grad-HS Dropout	-0.00003	0.00139	0.002	-0.00072	654	
	(0.00237)	(0.00124)	(0.00182)	(0.00098)		
HS Grad-HS Grad	0.0016+	0.00162*	0.0011	0.00045	2553	51
	(0.00096)	(0.00081)	(0.00106)	(0.00102)		
HS Grad-HS Dropout	0.00279	0.00113	0.00024	0.00033	1258	
	(0.00175)	(0.00097)	(0.00169)	(0.00032)		
HS Dropout-HS Dropout	-0.00246	-0.00104	0.00396	-0.0004	268	16
	(0.00399)	(0.00100)	(0.00309)	(0.00096)		
F-stat for Education Variables, Clustered at School-Pair Type Level					3.081	
F-pvalue					0.000	

Note: Each column and row displays a coefficient from the regression of an indicator for mutual friendship on a series of interaction terms between the scaled cohort composition variable of the column and the pair type variable of the row. All coefficients are standardized. A student pair sample of 6 million observations is estimated, which is composed of all unique pair of students within a school by grade by gender cell. The dependent variable is binary indicator of whether the two students in a pair both nominate each other as their friend. The independent variables are interactions between binary indictors for the type of a student pair and cohort composition variables. Maternal education type (defined by the match of the two parties' maternal education, as in each row, e.g. college grad-college grad) is interacted with percentage of college graduate mothers and the percentage of high school dropout mothers weighted by cohort size by gender. Racial type is interacted with percentage of black students, percentage of Hispanic students and percentage of Other/Asian students. This table presents coefficients from maternal education interactions. Coefficients of types with missing maternal education from one or both parties are omitted in this table because of ambiguous implication. All regressions control for school-gender-combined pair type fixed effect. The combined pair type is defined by both maternal education and race, e.g. white-high school grad-black-college grad. Standard errors are clustered at the school level. F-stat is for testing whether the 40 control variables (including missing maternal education) on maternal education jointly are significantly different from zero. It is calculated from the regression clustered by school-combined pair type because of the number of variables jointly tested (40) is large relative to the total number of control variables (100). **p<0.01, *p<0.05, and +p<0.1.

Table 5. Friendship Pattern Standardized Estimation from Matching Model for Race/Ethnicity

	Female			Male				
Racial Pair Type	% Black (1)	% Hispanic (2)	% Other (3)	% Black (4)	% Hispanic (5)	% Other (6)	Mean No. of Pairs in Cohort by Gender	Implied No. of Students
White-White	0.00515**	0.00272	-0.00001	0.00516**	0.0025+	0.00025	4890	70
	(0.00102)	(0.00174)	(0.00140)	(0.00170)	(0.00132)	(0.00082)		
White-Black	-0.00012	-0.00016	-0.00047	-0.00021	0.00100	0.00228**	1578	
	(0.00037)	(0.00071)	(0.00069)	(0.00049)	(0.00176)	(0.00057)		
White-Hispanic	-0.00125	-0.00181*	0.00134	0.00078	-0.00036	-0.00018	1643	
	(0.00273)	(0.00074)	(0.00097)	(0.00145)	(0.00053)	(0.00043)		
White-Other	-0.00325+	-0.00127	-0.00037	0.00249	0.00175	-0.00026	1246	
	(0.00169)	(0.00126)	(0.00046)	(0.00175)	(0.00152)	(0.00043)		
Black-Black	0.00114	-0.00261	0.00659**	0.00065	0.00583**	-0.0006	1223	35
	(0.00086)	(0.00277)	(0.00188)	(0.00047)	(0.00197)	(0.00112)		
Black-Hispanic	0.00010	-0.00135	-0.0006	0.00012	-0.00016	-0.00007	1035	
	(0.00107)	(0.00106)	(0.00069)	(0.00057)	(0.00056)	(0.00049)		
Black-Other	0.00013	-0.00416*	0.0006	-0.00033	0.00057	0.00056	536	
	(0.00028)	(0.00206)	(0.00127)	(0.00028)	(0.00126)	(0.00046)		
Hispanic-Hispanic	-0.00455	0.0014	0.00153	0.00381	-0.00026	0.00387*	1562	40
	(0.00458)	(0.00106)	(0.00162)	(0.00276)	(0.00028)	(0.00176)		
Hispanic-Other	-0.00653*	0.00089	-0.00111	0.00191	-0.00026	0.00003	664	
	(0.00285)	(0.00094)	(0.00069)	(0.00203)	(0.00064)	(0.00032)		
Other-Other	0.02303**	0.00215	-0.00041	-0.00234	-0.00148	0.00127+	312	18
	(0.00753)	(0.00400)	(0.00105)	(0.00282)	(0.00293)	(0.00071)		
F-stat for Race Variables, Clustered at School-Pair Type Level							2.662	
F-pvalue							0.000	

Note: See notes of table 4. This table presents the coefficients of racial interactions from the same regression described in the note of table 4. Standard errors are clustered at the school level. F-stat is for testing the 60 control variables related to race jointly are significantly different from zero. It is calculated from the regression clustered by school-combined pair type. **p<0.01, *p<0.05, and +p<0.1.

Table 6. Average GPA by Gender, Maternal Education and Race

	Female GPA	Male GPA
Full Sample Mean	2.879	2.721
Std. Dev.	(0.779)	(0.824)
Mom HS Dropout	2.602	2.452
Mom HS Graduate	2.860	2.685
Mom College Graduate	3.152	2.997
Mom Edu Missing	2.698	2.522
White	2.988	2.830
Black	2.660	2.457
Hispanic	2.620	2.478
Other	3.049	2.843
Obs.	36916	35303

Note: The statistics here are un-weighted simple means (and standard deviation) over all students in the subsamples.

Table 7. First Stage Correlation of Actual and Predicted Number of Friends by Maternal Edu

	Actual No. of Friends			
	Female		Male	
Predicted No. of Friends	Mom College Graduate	Mom HS Dropout	Mom College Graduate	Mom HS Dropout
	(1)	(2)	(3)	(4)
Mom College Graduate	0.860**		0.644**	
	(0.097)		(0.116)	
Mom HS Dropout		0.880**		1.101**
		(0.111)		(0.071)
Obs.	42331	42331	42364	42364
R-sq	0.205	0.137	0.186	0.136
F_iv	79.282	62.411	30.817	239.457
Fpvalue	0.000	0.000	0.000	0.000

Note: Predicted number of friends is generated by summing up the predicted probability of being friends from estimation using the pair level match sample for each student over all potential friends of a certain type (e.g. maternal college education). All regressions control for school-gender-student type fixed effect, school-grade fixed effect, and Guryan type controls for school level friendship pattern omitting the student's contribution. Standard errors are clustered at the school level. **$p<0.01$ and *$p<0.05$.

Table 8. Balancing Test for Friend Choice Sorting with Student Demographic Characteristics

	Predicted No. of Friends			
	Female		Male	
	Mom College Graduate	Mom HS Dropout	Mom College Graduate	Mom HS Dropout
Independent Variables	(1)	(2)	(3)	(4)
Age	-0.00005	-0.00002	-0.00019	-0.00012
	(0.00050)	(0.00032)	(0.00027)	(0.00025)
No. of People in Household	0.00015	0.00008	0.00009	0.00000
	(0.00031)	(0.00024)	(0.00019)	(0.00016)
No. of School Kids in Household	-0.00015	0.00012	-0.00030	-0.00010
	(0.00036)	(0.00024)	(0.00022)	(0.00017)
Live with Both Parents	0.00096	-0.00027	-0.00050	-0.00042
	(0.00082)	(0.00051)	(0.00055)	(0.00038)
Live with Biological Parents	-0.00083	0.00051	0.00174	-0.00020
	(0.00170)	(0.00141)	(0.00118)	(0.00092)
Mother's Edu in Single Year	0.00014	-0.00010	0.00041	0.00025
	(0.00030)	(0.00032)	(0.00034)	(0.00035)
Mother Born in US	-0.00159	-0.00032	-0.00032	0.00104
	(0.00119)	(0.00097)	(0.00091)	(0.00070)
Born in US	0.00074	0.00016	-0.00089	-0.00134
	(0.00121)	(0.00096)	(0.00083)	(0.00093)
Adopted	-0.00149	-0.00160	0.00060	0.00016
	(0.00191)	(0.00155)	(0.00137)	(0.00097)
Health Condition at Birth	-0.00055	-0.00164	-0.00027	-0.00024
	(0.00220)	(0.00123)	(0.00178)	(0.00142)
Obs.	42,331	42,331	42,364	42,364
R-squared	0.973	0.957	0.980	0.937
Ftest	0.576	0.516	0.676	0.600
Fpvalue	0.832	0.877	0.745	0.812

Note: Each column displays a separate regression of the instrument variable--predicted number of mutual friends with college graduate mothers or with high school dropout mothers, on ten predetermined demographics variables and the interaction terms of the predetermined demographic variables and their indicators for non-missing values. All regressions control for school-gender-student type fixed effect and school-grade fixed effect. Guryan type control for school level predicted number of mutual friends omitting the student him/herself is included in all regressions. Standard errors are clustered at the school level. **p<0.01 and *p<0.05.

Table 9. Effect of Friends Pattern on Student's GPA

	Female			Male		
	OLS	IV1	IV2	OLS	IV1	IV2
	(1)	(2)	(3)	(4)	(5)	(6)
No. of Friends with	0.153**	0.212*		0.155**	-0.076	
Mom College Graduate	(0.010)	(0.086)		(0.012)	(0.214)	
No. of Friends with	-0.011		0.013	-0.030		0.072
Mom HS Dropout	(0.014)		(0.138)	(0.024)		(0.105)
Guryan Control		0.044			-0.099	
Mom College Graduate		(0.047)			(0.126)	
Guryan Control			0.069			0.040
Mom HS Dropout			(0.077)			(0.078)
Obs.	36916	36663	36663	35303	35042	35042
R-squared	0.247	0.042	0.027	0.233	0.009	0.024
Type*School FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade*School FE	Yes	Yes	Yes	Yes	Yes	Yes
Weak IV F-stat		80.762	70.490		27.936	226.437
Std. Dev. of Actual No. of Friends with College Grad Maternal Education		0.655			0.566	

Note: Each column displays a separate regression of GPA on number of mutual friends with college graduate mothers and/or with high school dropout mothers. In IV regression, numbers of mutual friends with maternal college education are instrumented with corresponding predicted number of friends with maternal college education. All regressions control for school-gender-student type fixed effect and school-grade fixed effect. IV regressions include Guryan type control for school level friendship pattern in both first and second stage. In the following tables, all IV regressions control for school-gender-student type fixed effect, school-grade fixed effect and Guryan controls. Standard errors are clustered at the school level. **p<0.01 and *p<0.05.

Table 10. Effect of Friends Pattern on Student's GPA by Subject

	Female GPA				Male GPA			
	Math	English	Science	History	Math	English	Science	History
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No. of Friends with	0.027	0.317**	0.380**	0.125	-0.302	-0.085	-0.105	0.138
Mom College Graduate	(0.133)	(0.123)	(0.120)	(0.103)	(0.284)	(0.279)	(0.299)	(0.241)
Obs.	34516	35499	32351	32074	33380	33985	31164	30879
R-squared	0.029	0.029	0.024	0.051	-0.033	0.014	0.016	0.045
No. of Friends with	0.124	0.190	-0.011	-0.176	-0.087	0.075	0.107	0.011
Mom HS Dropout	(0.168)	(0.179)	(0.181)	(0.192)	(0.158)	(0.133)	(0.151)	(0.145)
Obs.	34516	35499	32351	32074	33380	33985	31164	30879
R-squared	0.024	0.023	0.037	0.036	0.021	0.025	0.029	0.037

Note: See note for Table 9. The regressions are the same as described in the note for Table 9, except that the dependent variables here are gpa of specific subjects. However, each column represents results from two IV regressions, because we stack the results of model IV1 (top panel) and IV2 (bottom panel) in order to save space. For the same reason, we omit the coefficients of Guryan controls and fixed effect lables. Standard errors are clustered at the school level. **p<0.01 and **p<0.05.

Table 11. Mechanism Analysis

	Self-Evaluate	Environment	Mental	Trouble	Misbehave	Addict	Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female							
No. of Friends with	0.258*	0.184*	0.176+	-0.057	-0.189+	-0.077	0.025
Mom College Graduate	(0.114)	(0.092)	(0.105)	(0.086)	(0.109)	(0.086)	(0.137)
R-squared	0.012	0.042	0.018	0.020	0.007	0.035	0.014
No. of Friends with	0.013	0.253+	-0.024	0.011	0.061	0.124	-0.210
Mom HS Dropout	(0.165)	(0.144)	(0.175)	(0.139)	(0.157)	(0.140)	(0.171)
R-squared	0.020	0.024	0.013	0.017	0.014	0.031	0.009
Obs.	36838	36456	35878	40006	37565	39070	37399
Male							
No. of Friends with	0.037	-0.017	0.059	-0.119	-0.209	-0.493*	0.084
Mom College Graduate	(0.187)	(0.210)	(0.179)	(0.233)	(0.235)	(0.242)	(0.184)
R-squared	0.015	0.017	0.016	0.017	0.012	-0.013	0.015
No. of Friends with	-0.003	0.140	-0.011	-0.037	-0.048	-0.155	-0.144
Mom HS Dropout	(0.136)	(0.142)	(0.151)	(0.143)	(0.151)	(0.131)	(0.138)
R-squared	0.013	0.019	0.013	0.015	0.015	0.032	0.012
Obs.	34796	34382	33464	38285	35731	37375	35361

Note: Each column and row of coefficients displays a coefficient from a separate IV regression; therefore each column contains coefficients from four IV regressions stacked. The regressions are the same as described in the note for Table 9, except that the dependent variables. Dependent variables are constructed by factor analysis of students' report on own mental status, behavior, school and family environment (see Appendix Table 18A for reference). Standard errors are clustered at the school level. **p<0.01, *p<0.05, and +p<0.1.

Table 12. Female Sample IV Results of Maternal College Friends Interacted with Own Maternal Education

	GPA	Self-Evaluate	Environment	Mental	Trouble	Misbehave	Addict	Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own Mom College Grad *No. of Friends with Mom College Grad	0.281**	0.250+	0.191	0.274*	-0.132	-0.270+	-0.103	0.165
	(0.092)	(0.152)	(0.119)	(0.118)	(0.099)	(0.156)	(0.105)	(0.168)
Own Mom HS Graduate *No. of Friends with Mom College Grad	0.112	0.234+	0.153	0.053	0.059	-0.107	-0.105	-0.217
	(0.117)	(0.125)	(0.126)	(0.138)	(0.135)	(0.129)	(0.121)	(0.160)
Own Mom HS Dropout *No. of Friends with Mom College Grad	0.087	0.443*	0.043	0.024	-0.097	-0.038	0.018	0.039
	(0.222)	(0.213)	(0.303)	(0.242)	(0.235)	(0.227)	(0.226)	(0.305)
Obs.	36663	36838	36456	35878	40006	37565	39070	37399
R-squared	0.033	0.012	0.037	0.012	0.012	0.004	0.032	-0.005
Weak IV F-stat	27.101	14.003	29.693	16.143	29.480	25.727	26.941	12.678

Note: Each column displays the coefficients from a separate IV regression. Interaction terms of actual number of maternal college friends and four dummies for own maternal education separately are instrumented with predicted number of maternal college friends interacted with four dummies for own maternal education respectively. The coefficients for the interaction term with missing maternal education are not presented. Dependent variables except GPA are constructed by factor analysis of students' report on own mental status, behavior, school and family environment (see Appendix Table 18A for reference). All regressions control for school-gender-student type fixed effect and school-grade fixed effect. Guryan type controls for school level friendship pattern (also interacted with dummies for own maternal education) are included in both first and second stage. Standard errors are clustered at the school level. **p<0.01, *p<0.05, +p<0.1.

Appendix Tables

Table 1A. Percentage of Mutual Friends, Deviation from School Share by Maternal Education

		Friend's Maternal Education			
	No. of Friends	High School Dropout	High School Graduate	College Graduate	Missing
Own Maternal Education					
Female					
High School Dropout	0.888	0.072	0.023	-0.061	-0.028
High School Graduate	1.171	-0.028	0.037	0.019	-0.058
College Graduate	1.302	-0.065	0.011	0.099	-0.065
Missing	0.795	0.010	0.015	0.002	-0.019
Male					
High School Dropout	0.524	0.053	0.026	-0.095	-0.035
High School Graduate	0.732	-0.014	0.048	0.032	-0.081
College Graduate	0.866	-0.045	0.028	0.100	-0.102
Missing	0.469	-0.001	0.024	0.005	-0.025
Maternal Education Distribution	100%	10.7%	44.3%	25.4%	19.7%

Note: We calculate the friendship frequencies by maternal education categories for every school and subtract the school fraction of the each student type by maternal education (columns) in order to get a deviation from expected based on frequency. A weighted average of this across all schools with weights based on number of students by type (rows) will deliver the empirical level of within school homophily.

Table 2A. Percentage of Mutual Friends, Deviation from School Share by Race

	No. of Friends	Friend's Race White	Black	Hispanic	Other
Own Race					
Female					
White	1.297	0.142	-0.336	-0.074	-0.071
Black	0.884	-0.330	0.329	-0.260	-0.063
Hispanic	0.797	-0.055	-0.071	0.173	-0.050
Other	0.960	-0.008	-0.090	-0.086	0.162
Male					
White	0.840	0.136	-0.260	-0.074	-0.080
Black	0.453	-0.302	0.297	-0.110	-0.021
Hispanic	0.464	-0.021	-0.012	0.132	-0.048
Other	0.632	0.020	-0.074	-0.053	0.150
Race Distribution	100%	55.2%	17.1%	17.1%	10.7%

Note: We calculate the friendship frequencies by race/ethnicity categories for every school and subtract the school fraction of the each student type by race/ethnicity (columns) in order to get a deviation from expected based on frequency. A weighted averagee of this across all schools with weights based on number of students by type (rows) will deliver the empirical level of within school homophily.

Table 3A. Predicted vs. Actual Friendship Patterns

		Actual No. of Friends		Predicted No. of Friends	
	Obs.	Mean	Std. Dev.	Mean	Std. Dev.
Friends Charateristics					
Total	84695	0.897	1.146	0.889	0.584
Mom HS Dropout	84695	0.079	0.301	0.079	0.147
Mom College Graduate	84695	0.273	0.614	0.270	0.289
Mom Education Missing	84695	0.119	0.360	0.118	0.182
Black	84695	0.117	0.454	0.117	0.285
Hispanic	84695	0.107	0.393	0.106	0.209
Other	84695	0.083	0.343	0.083	0.231

Table 4A. Friendship Pattern Estimation from Matching Model for Maternal Education

	Female		Male	
	% College Grad	% HS Dropout	% College Grad	% HS Dropout
	(1)	(2)	(3)	(4)
Maternal Education Pair Type				
College Grad-College Grad	1.267**	2.437	0.311	0.468
	(0.322)	(2.788)	(0.249)	(1.629)
College Grad-HS Grad	1.370**	1.685	0.388	0.828
	(0.228)	(1.394)	(0.281)	(0.717)
College Grad-HS Dropout	-0.011	0.877	0.738	-0.538
	(0.989)	(0.779)	(0.674)	(0.725)
HS Grad-HS Grad	0.669+	1.021*	0.407	0.334
	(0.400)	(0.512)	(0.391)	(0.760)
HS Grad-HS Dropout	1.162	0.711	0.090	0.245
	(0.732)	(0.612)	(0.626)	(0.240)
HS Dropout-HS Dropout	-1.027	-0.655	1.463	-0.299
	(1.665)	(0.630)	(1.143)	(0.711)
Std. Dev. of Cohort Composition/Cohort Size	0.0024	0.0016	0.0027	0.0013

Note: Each column and row displays a coefficient from the regression of an indicator for mutual friendship on a series of interaction terms between the scaled cohort composition variable of the column and the pair type variable of the row. A student pair sample of 6 million observations is estimated, which is composed of all unique pair of students within a school by grade by gender cell. The dependent variable is binary indicator of whether the two students in a pair both nominate each other as their friend. The independent variables are interactions between binary indictors for the type of a student pair and cohort composition variables. Maternal education type (defined by the match of the two parties' maternal education, as in each row, e.g. college grad-college grad) is interacted with percentage of college graduate mothers and the percentage of high school dropout mothers weighted by cohort size by gender. Racial type is interacted with percentage of black students, percentage of Hispanic students and percentage of Other/Asian students. This table presents coefficients from maternal education interactions. Coefficients of types with missing maternal education from one or both parties are omitted in this table because of ambiguous implication. All regressions control for school-gender-combined pair type fixed effect. The combined pair type is defined by both maternal education and race, e.g. white-high school grad-black-college grad. Standard errors are clustered at the school level. **p<0.01, *p<0.05, and +p<0.1.

Table 5A. Friendship Pattern Estimation from Matching Model for Race/Ethnicity

	Female			Male		
	%Black	%Hispanic	%Other	%Black	%Hispanic	%Other
	(1)	(2)	(3)	(4)	(5)	(6)
Racial Pair Type						
White-White	2.118**	1.459	-0.006	2.372**	1.327+	0.277
	(0.420)	(0.935)	(1.577)	(0.781)	(0.703)	(0.908)
White-Black	-0.049	-0.085	-0.525	-0.096	0.533	2.533**
	(0.154)	(0.383)	(0.775)	(0.224)	(0.938)	(0.638)
White-Hispanic	-0.515	-0.969*	1.513	0.357	-0.191	-0.197
	(1.123)	(0.396)	(1.093)	(0.667)	(0.281)	(0.476)
White-Other	-1.337+	-0.684	-0.417	1.145	0.930	-0.287
	(0.695)	(0.676)	(0.517)	(0.805)	(0.809)	(0.482)
Black-Black	0.469	-1.400	7.434**	0.298	3.099**	-0.669
	(0.355)	(1.487)	(2.117)	(0.217)	(1.049)	(1.240)
Black-Hispanic	0.042	-0.722	-0.677	0.056	-0.087	-0.074
	(0.439)	(0.568)	(0.780)	(0.264)	(0.295)	(0.540)
Black-Other	0.055	-2.233*	0.673	-0.152	0.305	0.627
	(0.117)	(1.107)	(1.434)	(0.128)	(0.672)	(0.506)
Hispanic-Hispanic	-1.871	0.751	1.729	1.752	-0.140	4.294*
	(1.883)	(0.569)	(1.832)	(1.267)	(0.151)	(1.959)
Hispanic-Other	-2.686*	0.475	-1.253	0.876	-0.137	0.038
	(1.171)	(0.506)	(0.782)	(0.933)	(0.342)	(0.354)
Other-Other	9.469**	1.155	-0.458	-1.077	-0.789	1.408+
	(3.097)	(2.148)	(1.187)	(1.297)	(1.556)	(0.789)
Std. Dev. of Cohort Composition/Cohort Size	0.0024	0.0019	0.0009	0.0022	0.0019	0.0009

Note: See notes of table 4A. This table presents the coefficients of racial interactions from the same regression described in the note of table 4A. Standard errors are clustered at the school level. **p<0.01, *p<0.05, and +p<0.1.

Table 6A. Frequency of Actual Tie

	Female	Male		Female	Male
Racial Pair			Mom Edu Pair		
White-White	0.012	0.007	College-College	0.013	0.008
White-Black	0.002	0.001	College-HS Grad	0.009	0.006
White-Hispanic	0.005	0.003	College-HS Dropout	0.004	0.003
White-Other	0.007	0.004	HS Grad-HS Grad	0.009	0.006
Black-Black	0.012	0.007	HS Grad-HS Dropout	0.006	0.004
Black-Hispanic	0.003	0.002	HS Dropout-HS Dropout	0.008	0.004
Black-Other	0.003	0.002			
Hispanic-Hispanic	0.005	0.003			
Hispanic-Other	0.003	0.002			
Other-Other	0.014	0.008			

Note: The frequency is calculated based on the student pair sample of 12 million observations. The numbers show the share of actual mutual ties of a particular type among all potential pairs by gender.

Table 7A. Heterogeneous Friendship Pattern Estimation from Matching Model for Maternal Education by Percent of Minority Subsamples

	Female			Male		
	% College Grad			% College Grad		
	High Minority	Low Minority	Diff T-stat	High Minority	Low Minority	Diff T-stat
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Edu Pair Type						
College Grad-	0.916+	1.464**		-0.046	0.771	
College Grad	(0.472)	(0.551)	-0.755	(0.256)	(0.473)	-1.519
College Grad-HS Grad	0.409	1.720**		-0.021	1.013	
	(0.334)	(0.264)	-3.079**	(0.177)	(0.619)	-1.606
College Grad-HS Dropout	-0.027	0.024		0.861	0.561	
	(0.683)	(2.123)	-0.023	(0.971)	(1.021)	0.213
HS Grad-HS Grad	0.096	0.779		0.786	0.225	
	(0.532)	(0.506)	-0.930	(0.483)	(0.475)	0.828
HS Grad-HS Dropout	2.057**	0.163		0.633	-0.586	
	(0.527)	(1.199)	1.446	(1.106)	(0.615)	0.963
HS Dropout-HS Dropout	-0.819	-2.242		1.186	1.666	
	(1.952)	(2.883)	0.409	(1.518)	(2.001)	-0.191
	% HS Dropout			% HS Dropout		
College Grad-	-2.272	6.761+		1.096	-0.119	
College Grad	(2.952)	(3.699)	-1.909+	(2.668)	(2.071)	0.360
College Grad-HS Grad	0.741	2.159		-0.012	1.206	
	(0.906)	(2.609)	-0.513	(1.150)	(0.886)	-0.839
College Grad-HS Dropout	-0.154	4.706+		-1.113	-0.056	
	(0.761)	(2.668)	-1.752+	(1.099)	(0.619)	-0.838
HS Grad-HS Grad	0.628	1.041		0.212	0.156	
	(0.761)	(0.662)	-0.409	(0.657)	(1.015)	0.046
HS Grad-HS Dropout	0.341	0.262		-0.025	-0.119	
	(0.584)	(1.402)	0.052	(0.395)	(0.448)	0.157
HS Dropout-HS Dropout	-1.314*	3.477		-0.572	-0.240	
	(0.573)	(3.415)	-1.384	(0.913)	(1.417)	-0.197

Note: We split all schools into two groups based share of minority students at school level in the way that the high minority group (no. of schools=68; no. of students=45,724) and the low minority group have approximately the same number of students (no. of schools=76; no. of students=44,216). We then estimated the two pair samples derived from these two groups of schools respectively. The model estimated is exact the same as described in note for Table 4, 5, 4A and 5A. Standard errors are clustered at the school level. F-stat for the 40 maternal education variables clustered by school-combined pair type is 1.545 (P=0.015; N=3,478,364) for the high minority subsample, and is 5.096 (P=0.000; N=2,892,751) for low minority subsample. Columns (1), (2), (4) and (5) are coefficients of education variables. Column (3) and (6) are the t-statistics for testing whether the coefficients from high and low minority subsamples are significant different from each other. **p<0.01, *p<0.05, and +p<0.1.

Table 8A. Heterogeneous Friendship Pattern Estimation from Matching Model for Maternal Education by Percent of Maternal College Subsamples

	Female			Male		
	% College Grad			% College Grad		
	High College	Low College	Diff T-stat	High College	Low College	Diff T-stat
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Edu Pair Type						
College Grad-College Grad	1.156**	0.862		0.610	-0.179+	
	(0.393)	(0.617)	0.402	(0.428)	(0.102)	1.793+
College Grad-HS Grad	1.483**	0.918*		0.241	0.576	
	(0.283)	(0.360)	1.234	(0.321)	(0.589)	-0.499
College Grad-HS Dropout	-1.327	2.508+		1.488	-0.339	
	(1.675)	(1.268)	-1.825+	(0.942)	(0.803)	1.476
HS Grad-HS Grad	0.827	0.549		0.607	0.284	
	(0.532)	(0.638)	0.335	(0.401)	(0.598)	0.449
HS Grad-HS Dropout	1.424	1.063		-0.806	0.528	
	(1.593)	(0.793)	0.203	(0.653)	(0.816)	-1.276
HS Dropout-HS Dropout	2.410	-2.014		0.660	1.289	
	(1.837)	(1.820)	1.711+	(2.395)	(1.343)	-0.229
	% HS Dropout			% HS Dropout		
College Grad-College Grad	3.641	1.591		-0.310	3.442*	
	(3.942)	(3.854)	0.372	(1.829)	(1.422)	-1.620
College Grad-HS Grad	-0.250	2.750		2.232	0.039	
	(1.602)	(1.865)	-1.220	(1.532)	(0.637)	1.322
College Grad-HS Dropout	0.863	0.695		-2.069	0.021	
	(2.154)	(0.548)	0.076	(1.477)	(0.759)	-1.259
HS Grad-HS Grad	-0.321	1.413*		0.488	0.306	
	(1.259)	(0.544)	-1.264	(1.153)	(0.840)	0.128
HS Grad-HS Dropout	-1.815	0.839		-0.687	0.311	
	(1.687)	(0.669)	-1.462	(1.785)	(0.300)	-0.551
HS Dropout-HS Dropout	6.184*	-0.564		2.381	-0.266	
	(3.082)	(0.660)	2.141*	(3.855)	(0.760)	0.674

Note: See note for Table 7A. We split all schools into two groups based share of maternal college students at school level in the way that the high maternal college group (no. of schools=66; no. of students= 45,909) and the low maternal college group have approximately the same number of students (no. of schools=78; no. of students= 44,031). We then estimated the two pair samples derived from these two groups of schools respectively. The model estimated is exact the same as described in note for Table 4, 5, 4A and 5A. Standard errors are clustered at the school level. F-stat for the 40 maternal education variables clustered by school-combined pair type is 3.288 (P=0.000; N=3,454,034) for the high maternal college subsample, and is 1.806 (P=0.001; N=2,917,081) for low maternal college subsample. **p<0.01, *p<0.05, and +p<0.1.

Table 9A. Heterogeneous Friendship Pattern Estimation from Matching Model for Maternal Education by School Size Subsamples

	Female			Male		
	% College Grad			% College Grad		
	Large School	Small School	Diff T-stat	Large School	Small School	Diff T-stat
Maternal Edu Pair Type						
College Grad-	1.027**	1.794**		-0.047	0.444	
College Grad	(0.250)	(0.537)	-1.295	(0.241)	(0.348)	-1.160
College Grad-HS Grad	1.182**	1.473**		0.755+	0.368	
	(0.345)	(0.247)	-0.686	(0.389)	(0.334)	0.755
College Grad-HS Dropout	0.311	-0.018		1.739	0.251	
	(0.382)	(1.653)	0.194	(1.186)	(0.664)	1.095
HS Grad-HS Grad	1.715**	0.466		1.345*	0.246	
	(0.606)	(0.450)	1.655	(0.507)	(0.439)	1.639
HS Grad-HS Dropout	2.656**	0.754		-0.141	0.184	
	(0.809)	(0.894)	1.578	(0.859)	(0.809)	-0.275
HS Dropout-HS Dropout	3.330*	-2.294		-0.067	2.318	
	(1.498)	(1.667)	2.509*	(0.875)	(1.772)	-1.207
	% HS Dropout			% HS Dropout		
College Grad-	4.364	2.375		3.659	0.175	
College Grad	(2.740)	(3.277)	0.466	(2.729)	(1.717)	1.081
College Grad-HS Grad	-0.360	1.935		0.406	0.886	
	(0.976)	(1.539)	-1.259	(0.891)	(0.788)	-0.404
College Grad-HS Dropout	0.474	1.176		-0.501	0.021	
	(0.574)	(0.930)	-0.642	(1.447)	(0.692)	-0.325
HS Grad-HS Grad	1.271	1.029+		0.022	0.361	
	(0.816)	(0.545)	0.247	(1.004)	(0.815)	-0.262
HS Grad-HS Dropout	0.670	0.683		1.229	0.246	
	(0.581)	(0.638)	-0.015	(0.828)	(0.270)	1.129
HS Dropout-HS Dropout	0.027	-0.801		1.242	-0.166	
	(0.754)	(0.781)	0.763	(0.967)	(1.012)	1.006

Note: See note for Table 7A. We split all schools into two groups based school size in the way that the large school group (no. of schools=38; no. of students= 45,890) and the small school group have approximately the same number of students (no. of schools=106; no. of students= 44,050). We then estimated the two pair samples derived from these two groups of schools respectively. The model estimated is exact the same as described in note for Table 4, 5, 4A and 5A. Standard errors are clustered at the school level. F-stat for the 40 maternal education variables clustered by school-combined pair type is 2.430 (P=0.000; N=4,509,465) for the large school subsample, and is 3.208 (P=0.000; N=1,861,650) for small school subsample. **p<0.01, *p<0.05, and +p<0.1.

Table 10A. Heterogeneous Friendship Pattern Estimation from Matching Model for Race/Ethnicity by Percent of Minority Subsamples

	Female			Male		
	% Black			% Black		
	High Minority	Low Minority	Diff T-stat	High Minority	Low Minority	Diff T-stat
Racial Pair Type						
White-White	2.183**	2.634**		2.830*	1.658+	
	(0.506)	(0.661)	-0.542	(1.294)	(0.856)	0.755
White-Black	0.274	0.135		-0.119	0.898	
	(0.242)	(0.343)	0.331	(0.250)	(1.628)	-0.617
White-Hispanic	-0.314	0.734		-0.171	4.555+	
	(1.152)	(1.651)	-0.521	(0.592)	(2.627)	-1.755+
White-Other	-1.655*	-0.384		1.092*	1.103	
	(0.677)	(1.487)	-0.778	(0.516)	(2.327)	-0.005
Black-Black	0.808*	-2.779		0.331	9.147	
	(0.345)	(3.666)	0.974	(0.227)	(15.376)	-0.573
Black-Hispanic	0.297	-16.553		0.133	3.134	
	(0.496)	(10.198)	1.650	(0.273)	(7.161)	-0.419
Black-Other	0.064	1.261		-0.069	-6.999**	
	(0.116)	(1.520)	-0.785	(0.069)	(2.026)	3.419**
Hispanic-Hispanic	-2.035	-8.676		2.044	-10.325	
	(1.899)	(10.113)	0.645	(1.269)	(10.551)	1.164
Hispanic-Other	-2.758*	0.795		0.881	-1.953	
	(1.216)	(5.421)	-0.640	(0.877)	(3.744)	0.737
Other-Other	9.606**	12.252*		-1.164	-1.943	
	(3.009)	(5.508)	-0.422	(1.394)	(4.506)	0.165
	% Hispanic			% Hispanic		
White-White	3.370**	0.790		1.898	0.982	
	(1.075)	(1.133)	1.652	(1.431)	(0.832)	0.553
White-Black	0.319	-0.672		1.104	-1.765	
	(0.346)	(1.709)	0.568	(0.970)	(1.492)	1.612
White-Hispanic	-0.144	1.450		0.089	-0.292	
	(0.338)	(3.306)	-0.480	(0.440)	(0.471)	0.591
White-Other	0.398	-1.504		3.004*	-0.260	
	(0.510)	(1.346)	1.321	(1.226)	(0.460)	2.493*
Black-Black	-1.389	31.537		3.344**	-5.347	
	(1.547)	(22.044)	-1.490	(1.091)	(18.540)	0.468
Black-Hispanic	-0.573	-22.301		0.186	6.102	
	(0.520)	(14.377)	1.510	(0.300)	(7.909)	-0.747
Black-Other	-1.143	-9.411		0.416	-2.238	
	(0.808)	(12.783)	0.646	(0.719)	(3.199)	0.809

Table Continued

	High Minority	Low Minority	Diff T-stat	High Minority	Low Minority	Diff T-stat
Hispanic-Hispanic	1.112*	5.079*		0.064	-2.805	
	(0.554)	(2.250)	-1.712+	(0.132)	(2.366)	1.211
Hispanic-Other	1.066+	1.464		0.314	-0.228	
	(0.539)	(1.084)	-0.329	(0.350)	(0.334)	1.120
Other-Other	3.664	-2.021		-0.997	0.815	
	(2.515)	(1.924)	1.795	(1.744)	(1.705)	-0.743
		% Other			% Other	
White-White	0.749	-0.111		1.161	-0.451	
	(1.280)	(1.938)	0.370	(1.728)	(1.095)	0.788
White-Black	-0.191	0.673		2.820**	1.551	
	(0.924)	(0.760)	-0.722	(0.840)	(1.252)	0.842
White-Hispanic	0.675	3.099		-0.683	1.081	
	(0.916)	(2.493)	-0.913	(0.521)	(1.292)	-1.266
White-Other	0.289	-0.446		-0.987	-0.326	
	(2.095)	(0.487)	0.342	(0.664)	(0.595)	-0.741
Black-Black	6.083**	30.839**		0.088	-13.059	
	(2.029)	(7.977)	-3.008**	(1.173)	(19.997)	0.656
Black-Hispanic	0.542	-16.591**		-0.227	-0.744	
	(0.757)	(4.688)	3.608**	(0.520)	(3.264)	0.156
Black-Other	1.580	0.177		0.783*	2.968	
	(1.755)	(5.015)	0.264	(0.351)	(2.094)	-1.029
Hispanic-Hispanic	2.741	-2.175		4.859*	1.582+	
	(2.217)	(1.696)	1.761+	(2.184)	(0.829)	1.403
Hispanic-Other	-0.885	-0.464		0.098	0.252	
	(1.079)	(0.556)	-0.347	(0.325)	(0.386)	-0.305
Other-Other	-3.069	0.644*		1.905	1.088**	
	(4.816)	(0.268)	-0.770	(1.536)	(0.329)	0.520

Note: This table presents the coefficients of racial/ethnic variables from the same regression described in Table 7A. Also see the note for the Table 7A. Standard errors are clustered at the school level. F-stat for the 60 racial/ethnic variables clustered by school-combined pair type is 2.692 (P=0.000; N=3,478,364) for the high minority subsample, and is 1.681 (P=0.001; N=2,892,751) for low minority subsample. Column (3) and (6) are the t-statistics for testing whether the coefficients from high and low minority subsamples are significant different from each other. **p<0.01, *p<0.05, and +p<0.1.

Table 11A. Heterogeneous Friendship Pattern Estimation from Matching Model for Race/Ethnicity by Percent of Maternal College Subsamples

	Female			Male		
	% Black			% Black		
	High College	Low College	Diff T-stat	High College	Low College	Diff T-stat
Racial Pair Type						
White-White	2.637**	1.818**		4.896*	1.607*	
	(0.778)	(0.529)	0.871	(2.412)	(0.675)	1.313
White-Black	-0.305	-0.000		0.866	-0.282+	
	(0.334)	(0.202)	-0.781	(0.832)	(0.165)	1.353
White-Hispanic	-1.788+	0.623		2.527	-0.371	
	(1.043)	(1.409)	-1.375	(1.573)	(0.713)	1.678
White-Other	-0.631	-1.967*		1.349	0.808	
	(1.166)	(0.802)	0.944	(1.583)	(0.777)	0.307
Black-Black	0.915	0.232		0.272	0.301	
	(0.597)	(0.493)	0.882	(0.428)	(0.244)	-0.059
Black-Hispanic	0.741	-0.214		-0.132	0.306	
	(0.500)	(0.597)	1.226	(0.498)	(0.390)	-0.692
Black-Other	-0.565	0.099		-0.784*	-0.019	
	(0.711)	(0.146)	-0.915	(0.320)	(0.048)	-2.364*
Hispanic-Hispanic	-5.886	-0.636		5.694	0.384	
	(5.248)	(1.782)	-0.947	(3.431)	(1.487)	1.420
Hispanic-Other	-1.406+	-4.709**		2.163	-0.427	
	(0.802)	(1.506)	1.936+	(1.810)	(0.674)	1.341
Other-Other	8.885**	6.953		-2.013	-0.298	
	(2.960)	(5.211)	0.322	(2.159)	(1.541)	-0.647
	% Hispanic			% Hispanic		
White-White	2.889*	0.025		1.688*	0.899	
	(1.122)	(1.322)	1.652	(0.662)	(1.186)	0.581
White-Black	0.076	-0.084		1.195	0.272	
	(0.626)	(0.621)	0.181	(2.137)	(0.481)	0.421
White-Hispanic	2.186	-0.939*		-0.315	-0.241	
	(1.413)	(0.393)	2.131*	(0.715)	(0.343)	-0.093
White-Other	-0.227	-1.193		0.806	1.012	
	(0.884)	(1.003)	0.723	(0.888)	(1.116)	-0.144
Black-Black	-4.340*	1.859		2.970*	3.332*	
	(2.154)	(1.313)	-2.457*	(1.379)	(1.453)	-0.181
Black-Hispanic	1.046	-0.591		-0.545	0.118	
	(1.506)	(0.517)	1.028	(0.772)	(0.379)	-0.771
Black-Other	-1.148	-2.916+		0.394	-0.421	
	(1.119)	(1.559)	0.921	(0.897)	(1.047)	0.591

Table Continued

	High College	Low College	Diff T-stat	High College	Low College	Diff T-stat
Hispanic-Hispanic	1.190	0.808		-0.061	0.045	
	(1.492)	(0.575)	0.239	(0.172)	(0.362)	-0.264
Hispanic-Other	0.748	1.313*		0.621	0.371+	
	(0.729)	(0.622)	-0.590	(0.744)	(0.199)	0.325
Other-Other	-2.185	5.191**		1.920	-2.951**	
	(2.363)	(1.577)	-2.596**	(1.687)	(0.379)	2.817**
		% Other			% Other	
White-White	1.768	-4.037+		-0.606	1.205	
	(1.642)	(2.225)	2.099*	(0.851)	(1.836)	-0.895
White-Black	-0.562	-0.183		2.361**	0.344	
	(0.841)	(1.158)	-0.265	(0.617)	(0.816)	1.972*
White-Hispanic	0.613	1.740		-0.337	-0.654	
	(1.362)	(1.710)	-0.516	(0.621)	(1.372)	0.210
White-Other	-0.612	-0.497		-0.246	0.129	
	(0.638)	(0.706)	-0.121	(0.406)	(1.582)	-0.230
Black-Black	10.744**	0.976		-0.507	-2.099	
	(3.296)	(3.488)	2.035*	(1.627)	(3.185)	0.445
Black-Hispanic	-1.251	-1.099		-0.153	0.462	
	(0.958)	(2.576)	-0.055	(0.664)	(1.394)	-0.398
Black-Other	0.932	-1.778		1.203*	0.392	
	(1.683)	(3.455)	0.705	(0.455)	(1.132)	0.665
Hispanic-Hispanic	2.362	-0.149		6.031*	0.157	
	(2.496)	(2.341)	0.734	(2.586)	(1.273)	2.038*
Hispanic-Other	-0.320	-3.048**		-0.288	1.535	
	(0.672)	(0.943)	2.356*	(0.321)	(1.997)	-0.901
Other-Other	0.226	-2.345		0.543	7.491	
	(1.124)	(3.768)	0.654	(0.869)	(5.816)	-1.182

Note: This table presents the coefficients of racial/ethnic variables from the same regression described in Table 8A. Also see the note for the Table 8A. Standard errors are clustered at the school level. F-stat for the 60 racial/ethnic variables clustered by school-combined pair type is 1.985 (P=0.000; N=3,454,034) for the high maternal college subsample, and is 2.091 (P=0.000; N=2,917,081) for low maternal college subsample. **p<0.01, *p<0.05, and +p<0.1.

Table 12A. Heterogeneous Friendship Pattern Estimation from Matching Model for Race/Ethnicity by School Size Subsamples

	Female			Male		
	% Black			% Black		
	Large School	Small School	Diff T-stat	Large School	Small School	Diff T-stat
Racial Pair Type						
White-White	0.597	2.256**		0.958	2.477**	
	(1.565)	(0.437)	-1.021	(2.673)	(0.820)	-0.543
White-Black	-1.085**	0.093		-1.299**	0.002	
	(0.381)	(0.167)	-2.832**	(0.298)	(0.232)	-3.445**
White-Hispanic	0.764	-0.726		0.150	0.327	
	(0.890)	(1.359)	0.917	(1.112)	(0.800)	-0.129
White-Other	-0.010	-1.576+		1.301	0.962	
	(0.970)	(0.808)	1.240	(1.404)	(0.917)	0.202
Black-Black	1.483	0.423		0.097	0.352	
	(0.897)	(0.396)	1.081	(0.703)	(0.245)	-0.343
Black-Hispanic	-0.154	0.134		-0.322	0.056	
	(0.627)	(0.526)	-0.352	(0.527)	(0.333)	-0.606
Black-Other	-0.166	0.066		-1.225	-0.113	
	(1.307)	(0.121)	-0.177	(0.989)	(0.107)	-1.118
Hispanic-Hispanic	3.064*	-4.615*		3.381	1.234	
	(1.462)	(2.159)	2.945**	(2.918)	(1.409)	0.663
Hispanic-Other	-0.931	-3.646*		3.850*	-0.017	
	(1.132)	(1.422)	1.494	(1.796)	(0.616)	2.037*
Other-Other	4.802*	11.068**		-3.675	-1.687	
	(1.917)	(3.594)	-1.538	(4.504)	(1.710)	-0.413
	% Hispanic			% Hispanic		
White-White	-2.263+	1.857+		2.495**	1.148	
	(1.272)	(1.008)	-2.539*	(0.875)	(0.794)	1.140
White-Black	-0.355	0.372		-1.023	0.939	
	(0.507)	(0.484)	-1.037	(0.849)	(1.131)	-1.387
White-Hispanic	-0.372	-1.094**		-0.907+	-0.184	
	(0.470)	(0.414)	1.153	(0.516)	(0.305)	-1.206
White-Other	-0.210	-0.701		0.068	1.038	
	(0.885)	(0.835)	0.404	(0.631)	(0.934)	-0.861
Black-Black	1.434	-1.917		1.881	3.065**	
	(2.360)	(1.764)	1.137	(1.509)	(1.143)	-0.625
Black-Hispanic	-0.668	-0.803		-0.538	-0.132	
	(0.467)	(0.691)	0.162	(0.364)	(0.498)	-0.658
Black-Other	-0.700	-2.343+		0.083	0.220	
	(2.205)	(1.407)	0.628	(1.055)	(0.864)	-0.100

Table Continued

	Large School	Small School	Diff T-stat	Large School	Small School	Diff T-stat
Hispanic-Hispanic	-0.756*	1.180+		-0.997*	-0.106	
	(0.314)	(0.597)	-2.870**	(0.461)	(0.182)	-1.798
Hispanic-Other	-0.155	0.951		-0.748+	-0.240	
	(0.516)	(0.736)	-1.230	(0.401)	(0.483)	-0.809
Other-Other	-0.661	2.239		2.451	-1.446	
	(1.092)	(2.484)	-1.069	(2.344)	(1.630)	1.365
		% Other			% Other	
White-White	3.284**	-0.301		0.230	-0.031	
	(0.761)	(1.914)	1.741	(1.044)	(1.318)	0.155
White-Black	-1.193*	1.473*		2.556**	3.000**	
	(0.586)	(0.676)	-2.980**	(0.615)	(1.064)	-0.361
White-Hispanic	-1.614	3.882**		-0.295	-0.098	
	(1.041)	(1.381)	-3.178**	(1.035)	(0.529)	-0.169
White-Other	-1.430**	-0.153		0.399	-0.728	
	(0.382)	(0.738)	-1.537	(0.512)	(0.487)	1.595
Black-Black	5.630**	7.776+		1.165	-1.243	
	(1.638)	(4.512)	-0.447	(1.131)	(1.583)	1.238
Black-Hispanic	-1.342+	-1.107		-1.055+	1.836*	
	(0.763)	(1.950)	-0.112	(0.559)	(0.843)	-2.858**
Black-Other	-1.522	3.484+		0.430	0.236	
	(0.946)	(1.833)	-2.427*	(0.490)	(0.766)	0.213
Hispanic-Hispanic	1.674	0.452		4.673*	3.056	
	(1.970)	(3.464)	0.307	(2.232)	(3.771)	0.369
Hispanic-Other	-1.107+	-2.075		-0.368	0.028	
	(0.605)	(1.408)	0.632	(0.350)	(0.495)	-0.653
Other-Other	-0.246	-0.930		1.303*	0.839	
	(0.539)	(3.676)	0.184	(0.580)	(1.046)	0.388

Note: This table presents the coefficients of racial/ethnic variables from the same regression described in Table 9A. Also see the note for the Table 9A. Standard errors are clustered at the school level. F-stat for the 60 racial/ethnic variables clustered by school-combined pair type is 2.606 (P=0.000; N=4,509,465) for the large school subsample, and is 2.442 (P=0.000; N=1,861,650) for the small school subsample. **p<0.01, *p<0.05, and +p<0.1.

Table 13A. Statistics Summary

Variables	Obs.	Mean	Std. Dev.	Min	Max
Male	84695	0.500	0.500	0	1
Age	84406	14.976	1.716	10	19
White	84695	0.552	0.497	0	1
Black	84695	0.171	0.376	0	1
Hispanic	84695	0.171	0.376	0	1
Other	84695	0.107	0.309	0	1
Mom High School Dropout	84695	0.107	0.309	0	1
Mom High School Graduate	84695	0.443	0.497	0	1
Mom College Graduate	84695	0.254	0.435	0	1
Mom Education Missing	84695	0.197	0.397	0	1
School Size	84695	998.455	585.667	44	2551
Cohort Size by Gender	84695	142.012	79.411	2	394
GPA	72219	2.802	0.805	1	4
Nominating Any Out-of-Grade within School Friend	84695	0.260	0.439	0	1
No. of People in Household	81537	4.289	1.144	1	6
No. of School Kids in Household	79063	0.711	0.929	0	6
Mother Born in US	74131	0.824	0.381	0	1
Father Born in US	60928	0.822	0.382	0	1
Born in US	82038	0.905	0.293	0	1
Live with Both Parents	80930	0.728	0.445	0	1
Live with Mother	82027	0.920	0.272	0	1
Live with Father	81730	0.762	0.426	0	1
Mother’s Edu in Single Year	68042	13.365	2.369	0	17
Father’s Edu in Single Year	55115	13.598	2.519	0	17
Adopt	82016	0.028	0.164	0	1
Health Condition at Birth	75931	0.019	0.135	0	1
No. of Valid Female Friend Nomination	84695	2.446	2.037	0	5
No. of Valid Male Friend Nomination	84695	2.362	1.997	0	5

Note: The low numbers of observations of variables “mother’s edu in single year” and “father’s edu in single year” are caused by excluding the students who report their parents’ education as “I don’t know” or missing.

Table 14A. First Stage Correlation of Actual and Predicted Number of Friends, Total and by Race

	Actual No. of Friends							
	Female				Male			
	Total	Black	Hispanic	Other	Total	Black	Hispanic	Other
Predicted No. of Friends	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	0.793**				0.886**			
	(0.086)				(0.092)			
Black		0.697**				0.637**		
		(0.122)				(0.104)		
Hispanic			0.812**				0.800**	
			(0.123)				(0.086)	
Other				0.845**				0.847**
				(0.083)				(0.139)
N	42331	42331	42331	42331	42364	42364	42364	42364
R-sq	0.208	0.374	0.238	0.249	0.182	0.265	0.165	0.212
F_iv	84.800	32.512	43.525	103.191	93.658	37.835	85.824	37.131
Fpvalue	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Predicted number of friends is generated by summing up the predicted probability of being friends from estimation using the pair level match sample for each student over all potential friends of a certain type (e.g. black). All regressions control for school-gender-student type fixed effect, school-grade fixed effect, and Guryan type controls for school level friendship pattern omitting the student's contribution. Standard errors are clustered at the school level. **p<0.01 and *p<0.05.

Table 15A. Multivariate Robustness Test

	Female GPA			Male GPA		
	2IV	3IV	5IV	2IV	3IV	5IV
	(1)	(2)	(3)	(4)	(5)	(6)
No. of Friends with	0.210*	0.270**	0.272**	-0.073	-0.219	-0.244
Mom College Graduate	(0.085)	(0.099)	(0.100)	(0.213)	(0.250)	(0.246)
No. of Friends with	0.017	0.059	0.051	0.062	-0.149	-0.163
Mom HS Dropout	(0.130)	(0.141)	(0.148)	(0.106)	(0.188)	(0.194)
Total No. of Friends		-0.064	-0.016		0.166	0.239
		(0.069)	(0.077)		(0.113)	(0.132)
No. of Black Friends			-0.135			-0.230
			(0.099)			(0.263)
No. of Hispanic Friends			-0.006			-0.132
			(0.148)			(0.226)
Obs.	36663	36663	36663	35042	35042	35042
R-squared	0.042	0.028	0.029	0.009	0.017	0.011
Weak IV F-stat	45.976	22.605	13.871	13.723	10.317	5.452

Note: Each column displays a separate regression of GPA on number of mutual friends in different categories. Numbers of mutual friends are instrumented with the corresponding predicted number of friends. All regressions control for school-gender-cross pair type fixed effect and school-grade fixed effect. Guryan type controls for school level friendship pattern are included for each instrumented variable in both first and second stage. Standard errors are clustered at the school level. **p<0.01 and *p<0.05.

Table 16A. Effect of Friends Pattern on GPA using IV from Match Sample Estimation by School Characteristics

	Female GPA			Male GPA		
	Minority Share Split	Maternal College Split	School Size Split	Minority Share Split	Maternal College Split	School Size Split
	(1)	(2)	(3)	(4)	(5)	(6)
No. of Friends with	0.243**	0.229*	0.227**	-0.072	-0.071	-0.024
Mom College Graduate	(0.084)	(0.092)	(0.081)	(0.205)	(0.198)	(0.228)
Obs.	36663	36663	36663	35042	35042	35042
R-squared	0.039	0.041	0.041	0.010	0.010	0.020
Weak IV F-stat	59.600	68.543	62.979	35.497	24.593	30.840
Fpvalue	0.000	0.000	0.000	0.000	0.000	0.000
No. of Friends with	-0.011	0.012	0.039	0.044	0.108	0.052
Mom HS Dropout	(0.135)	(0.134)	(0.128)	(0.104)	(0.108)	(0.106)
Obs.	36663	36663	36663	35042	35042	35042
R-squared	0.027	0.027	0.027	0.024	0.023	0.024
Weak IV F-stat	98.063	85.614	106.974	241.296	186.924	242.110
Fpvalue	0.000	0.000	0.000	0.000	0.000	0.000

Note: Each column and row of coefficients displays a coefficient from a separate IV regression; therefore each column contains coefficients from two IV regressions stacked (IV1 in the top panel with no. of maternal college friends instrumented and IV2 in the bottom panel with no. of maternal high school dropout friends instrumented). The instruments are derived from estimations using stratified pair level match subsamples by share of minority, share of maternal college students, or school size respectively, as presented in Table 7A-12A. Also see the notes of Table 7A-12A. Standard errors are clustered at the school level. All regressions control for school-gender-student type fixed effect and school-grade fixed effect. IV regressions include Guryan type control for school level friendship pattern in both first and second stage. **p<0.01 and *p<0.05.

Table 17A. Heterogeneous Effect of Friends Pattern on GPA among Female using IV by School Characteristics

Dependent Var: GPA	High Minority (1)	Low Minority (2)	High College (3)	Low College (4)	Large School (5)	Small School (6)
No. of Friends with	0.192	0.324**	0.129	0.418*	0.113	0.304**
Mom College Graduate	(0.115)	(0.109)	(0.089)	(0.198)	(0.111)	(0.111)
Obs.	18170	18493	18502	18161	18501	18162
R-squared	0.043	0.021	0.046	0.022	0.033	0.044
Weak IV F-stat	48.249	20.077	51.313	19.950	20.066	43.889
Fpvalue	0.000	0.000	0.000	0.000	0.000	0.000
No. of Friends with	-0.138	0.191	-0.039	0.035	0.034	0.040
Mom HS Dropout	(0.167)	(0.224)	(0.204)	(0.169)	(0.256)	(0.145)
Obs.	18170	18493	18502	18161	18501	18162
R-squared	0.024	0.014	0.026	0.029	0.018	0.036
Weak IV F-stat	51.935	71.762	53.343	50.138	124.581	67.353
Fpvalue	0.000	0.000	0.000	0.000	0.000	0.000

Note: Each column and row of coefficients displays a coefficient from a separate IV regression; therefore each column contains coefficients from two IV regressions stacked (IV1 in the top panel with no. of maternal college friends instrumented and IV2 in the bottom panel with no. of maternal high school dropout friends instrumented). The instruments are derived from estimations using stratified pair level match subsamples by share of minority, share of maternal college students, or school size respectively, as presented in Table 7A-12A. Also see the notes of Table 7A-12A. We estimate the coefficients within subsamples at the student level for female. Standard errors are clustered at the school level. All regressions control for school-gender-student type fixed effect and school-grade fixed effect. IV regressions include Guryan type control for school level friendship pattern in both first and second stage. **p<0.01 and *p<0.05.

Table 18A. Factor Analysis Elements

	Survey Questions
Self Evaluation	How strong do you agree or disagree with each of the following statements? --I am physically fit. --I have a lot to be proud of. --I like myself just the way I am. --I feel like I am doing everything just right. --I have a lot of good qualities. In general, how hard do you try to do your school work well?
Environment Evaluation	How strong do you agree or disagree with each of the following statements? --I feel close to people at this school. --I feel like I am part of this school. --The students at this school are prejudiced. --The teachers at this school treat students fairly. --I feel safe in my school. --I am happy to be at this school.
Mental Health	How often did you feel depressed or blue in the last month? How often did you afraid of things in the last month? How strong do you agree or disagree with each of the following statements? --I feel loved and wanted. --I feel socially accepted. What do you think are the chances you will be killed by age 21.
Trouble at School	Since school started this year, how often have you had trouble: --getting along with your teachers? --paying attention in school? --getting your homework done? --getting along with other students?
Problematic Behavior	During the past twelve months, how often did you: --lie to your parents or guardians? --skip school without an excuse? In the past year, how often have you gotten into a physical fight?
Smoking and Drinking	During the past twelve months, --did you smoke cigarettes every week? --did you drink beer, wine, or liquor every week? --did you get drunk every week? Have you had a drink of beer, wine, or liquor—not just a sip or a taste of someone else's drink—more than two or three times in your life?
Health Status	In general, how is your health? How strongly do you agree or disagree with each of the following statements? --I seldom get sick. --When I do get sick, I get better quickly. In the last month, how often did a health or emotional problem cause you to: --miss a day of school? --miss a social or recreational activity?

Note: all variables from original dataset are converted to binary indicators to simplify the factor analysis.