Friendship Patterns in Schools: An Analysis Using the National Longitudinal Study of Adolescent Health

by

Anil Sathia Nathan

Department of Economics
Duke University

Date: ______________________________

Approved:

____
Peter Arcidiacono, Supervisor

____
Thomas Nechyba

____
Alessandro Tarozzi

____
Christopher Timmins

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2008

ABSTRACT

Friendship Patterns in Schools: An Analysis Using the National Longitudinal Study of Adolescent Health

by

Anil Sathia Nathan

Department of Economics
Duke University

Date: ______________________________
Approved:

____
Peter Arcidiacono, Supervisor

____
Thomas Nechyba

____
Alessandro Tarozzi

____
Christopher Timmins

An abstract of a dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2008

Abstract

The rigorous economic analysis of peer group formation is a burgeoning subject. Much has been written about how peers influence an individual's behavior, and these effects are quite prevalent. However, less has been written on how exactly these peer groups begin and the resulting consequences of their formation. A reason for the dearth of knowledge on peer group formation is the lack of quality data sets that clearly define one's peers. To resolve this issue, the first chapter of this document explores data which allows a peer group to be defined openly through self nominations. Using these nominations as well as characteristics of the students and their friends, it is possible to see on what dimensions these individuals are sorting into friendships. The data suggests that there is heavy sorting within race and academic ability. Additionally, tests for statistical discrimination on race and academics show that it is exhibited towards blacks and Hispanics. There is also weak evidence of statistical discrimination against whites. Empirical analysis also shows that the degree of statistical discrimination decreases for blacks and Hispanics over a year; however, there is little change for whites over the same period. This result suggests a process of learning about a noisy signal on academic characteristics. Once a peer group is formed, however, the effects of the peer group become much more interesting. The second chapter of this document attempts to find an effect of having friends of a similar race and who are involved in similar activities. Using a strategy that corrects for the endogeneity of peer effects by instrumenting using variables at the "grade within school" level, it is shown that friendship diversity can help whites increase achievement. Although not much significance was found with other races, most of the strategies pushed towards the direction of racial diversity aiding achievement. Regarding extracurricular activities, it is found that there is a benefit in having friends in common individual academic activities, conditional on the respondent only belonging to academic or scholastic clubs. There are insignificant effects in having friends in common sports, conditional on the respondent only participating in sports. Potential future work for both chapters can include models describing the benefit

of having various friends and the probability of forming those friendships, which can be used to simulate redistribution policies.

Contents

List of Tables

Acknowledgements

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis. I thank my dissertation committee, which consists of Peter Arcidiacono, Thomas Nechyba, Christopher Timmins, and Alessandro Tarozzi, the participants of the Applied Microeconomics Lunch Group at Duke University, and Paul Dudenhefer for their insightful comments and suggestions. I also would like to thank my father, sister, and lovely fiancée for their patience and loving support.

Chapter 1

Sorting and Statistical Discrimination in the National Longitudinal Study of Adolescent Health

1.1 Introduction

While the analysis of peer group formation and peer effects is well established in sociology and psychology, it has only recently been broached using rigorous economic analysis. Generally, the effect of one's peer group on one's behavior in many facets is very strong and profound. Knowing how one's peers actually come into being, however, is a subject that is not as well understood as what happens after the group is formed. Nevertheless, the process of peer group formation can be important in many ways.

Specifically, an understanding of sorting into peer groups is important whenever the distribution of characteristics on which individuals sort affects an outcome. Examples of such characteristics are race, academics, and attitudes (Clotfelter 2004). Two of the mechanisms that can affect sorting along these lines are the following.

- Homophily: A sociological term where individuals associate with others of similar characteristics (Kandel 1978).
- Statistical Discrimination: The perception and treatment of individuals based on the discernment of group characteristics; this is also known as stereotyping.

These two different but related processes may affect, for example, how students who are redistributed into schools based on race are received by the original members of the school, how the original members are received by those redistributed, and if any economic returns can be gained by the distributional changes. For example, Arcidiacono and Vigdor (2005) find that there are weak effects between racial diversity in college and post-graduation outcomes for white and Asian students. Perhaps the reason that the effects are not stronger are that white and Asian students remain entrenched in their racial groups due to preference (homophily) or that they only associate with members of other races of whom they perceive give a signal that is both of the following:

- Different from a mean perceived signal from the other races
- Similar to their own characteristics

Both of the above characteristics occurring together are hallmarks of statistical discrimination. If, for example, the signal is on intelligence/academics, both homophily and statistical discrimination may dampen any sort of gains to diversity that are attempted to be exploited by administrators and policy makers.

Knowing if homophily and statistical discrimination exist and the magnitudes of such effects based on a change in distributions is an important part of designing any sort of redistribution policy, such as school redistricting and affirmative action.[1] It is also important to know if repeated contact with peers over time results in any behavioral change with regards to homophily and statistical discrimination. With repeated exposure to signals and characteristics of potential peers, homophily and statistical discrimination may change. For example, after a school redistribution program is first implemented, homophily and statistical discrimination may decrease in magnitude. This result may further the goals of administrators and policy makers with regards to the policy's initial intentions of having more interracial contact; this may in turn allow for a better scholastic experience and better economic outcomes outside of school.

In order to address the above issues, this chapter will conduct the following analysis.

- Determine patterns of sorting across racial, academic, and attitudinal lines using two waves of the the National Longitudinal Study of Adolescent Health (Add Health).
- Develop and estimate a model of homophily.
- Alter the homophily model to develop and estimate a model of statistical discrimination.

[1]A good example is Boston's Metropolitan Council for Educational Opportunity (Metco) Program, which redistributes minorities that must meet a certain academic standard across schools in the Greater Boston area. The main aims of the program are to help desegregate Boston area schools as well as provide opportunities to certain minorities by transferring them to advantageous school districts (Drews 2006).

Section 1.2 provides an in-depth background of the research and issues involved in peer effects analysis, peer group formation, and policies of integration and redistribution, as well as explaining the value added by this chapter to the literature. Section 1.3 describes the Add Health data and explains the key features of Add Health that are exploited. Section 1.4 shows descriptive evidence of sorting along racial, academic, and attitudinal lines. Section 1.5 develops a model of homophily and shows the results that homophily does at least weakly exist along certain lines and changes over time. Section 1.6 alters the model of homophily to describe a measure of statistical discrimination and presents results which show that statistical discrimination exists towards blacks and Hispanics, which changes over time. Section 1.7 revisits the motivation of the chapter presented above with results in hand, and outlines implications and future work.

1.2 Literature Review

A large portion of the literature on peer groups focuses on their effects rather than their formation. Peer groups have been shown to be very important in many facets of life, and can influence behavior immensely. Although this chapter does not analyze any peer effects per se, it is important to know that many papers have attempted to analyze these effects. Case and Katz (1991) explain that there exist significant neighborhood peer effects on drug and alcohol usage, church attendance, and unemployment from youth through adulthood. Duncan, Boisjoly, Kremer, Levy, and Eccles (2006) show that the racial composition of freshman housing assignments has an effect on student attitudes towards their peers. Even the potential biases of peer groups are explored in depth, such as the reflection problem in Manski (1993) where reference (or peer) groups may exhibit an endogenous social effect on individuals, as well as potential corrections using non-linearities outlined in Brock and Durlauf (2001). Weinberg (2003) goes even further by suggesting a model of social interaction with endogenous association using the National Longitudinal Study of Adolescent Health (Add Health), the same set of data that is used in this chapter. In many of these papers, however, peer groups are very anomalous. Add Health has clearly defined peer groups, so there can be more confidence that the effects calculated are from actual peer groups instead of assumed peer groups. The next chapter of this document will explore actual peer effects.

Discerning how a peer group forms and on what dimensions they sort has not been attempted as much due to the lack of proper data on peer groups. However, the theory of group formation has been explored in detail, especially by psychologists. Raino (1966) and Tuma and Hallinan (1978) believe that similarity and status are two important precursors to friendship. Blau (1964) offers a model where an agent calculates the expected benefit and cost of forming a friendship before making a decision on the friend. Akerlof and Kranton (2002) form a theoretical framework of group formation amongst students. They suggest that students match their characteristics to a set of pre-existing social categories. Students receive greater utility by matching to a group that is most similar to their observed characteristics. After first choosing the group, they then choose how much effort they put into schooling (an example of a peer effect), which is conditional on group choice. However, the authors do not empirically test their premises. Marmaros and Sacerdote (2006) suggest a model where the expected benefit of a friendship is dependent on information gathered and any shared experiences, while the cost is the time used to develop the friendship. They do not assume that the individual can predict with a reasonable degree of certainty who would be a good friend. Arcidiacono, Khan, and Vigdor (2008) develop a model of interracial contact where individuals want to match with a friend who is similar academically, but where the signal of academic quality is noisy. This results in individuals statistically discriminating over any potential friends.

Marmaros and Sacerdote (2006) use a unique dataset from Dartmouth that measures the level of social interaction between any two individuals as the amount of e-mail sent between them. They find that the greatest dimensions of sorting are along racial lines and geographic boundaries by estimating poisson regressions of the the number of e-mails sent between any two people on various characteristics. Although e-mails may be a reasonable proxy for friendships, this chapter aims to use the more concrete friendship nomination data in Add Health. Foster (2005) and Arcidiacono, Khan, and Vigdor (2008) also use datasets that list characteristics of respondents and how many friends that they have across different lines.[2] However, it is not possible to identify the actual friendship nominations using these sets of data. Therefore, demographic data, apart from the line that is being matched, on friendship nominations is unavailable. Add Health has complete demographic data on friendship nominations within a particular school, so it is possible observe matches across multiple lines. This feature of Add Health is exploited, which is the true value added by this chapter.

Racial diversity is a very important topic that pertains to schools; programs such as school desegregation and busing have been implemented with the intention of forming new peer groups and fostering better educational and cultural outcomes. One such outcome is the elimination of the black/white achievement gap. Bowen and Bok (2001) argue that learning across races takes place and is quite useful; while Clotfelter (2004) chronicles how important a topic such as school desegregation in America is to both whites and blacks alike, and show how desegregation programs may have led to "white flight." On the other hand, Bifulco and Ladd (2007) show that black students choose charter schools in North Carolina with a higher portion of black peers, despite poorer results than private schools. In essence, racial homogeneity is chosen over academic excellence. However, there is some dispute as to the importance of racial composition on academic outcomes. Rumberger and Palardy (2001) argue that the socioeconomic level of students' schools as well as students' own socioeconomic status have about the same impact on achievement growth for both advantaged and disadvantaged students as well as for both white and non-white students. These findings question whether integration policies have any impact at all.

It is also argued that integration policies may actually lead to more segregation if there is a small minority present. Moody (2001) argues that segregation through clubs and sports can result in the appearance of segregation based on race. The same process could happen if students are on academic "tracks," such as honors classes. In these cases, racial sorting

[2] Foster (2005) uses data from the University of Maryland registrar, while Arcidiacono, Khan, and Vigdor (2008) use the College and Beyond data.

patterns like homophily and behaviors such as statistical discrimination can be confounded through other factors. However, Xie and Zeng (2002) model the selection of friendship based on "choice" (which includes dimensions such as race) and "opportunity" (which includes scholastic institutions that segregate) using a conditional logit framework. They find that race is the most important factor in choosing friendships. Regardless, this chapter checks for robustness of results by controlling for these issues through a random effects framework, with individuals as the group variable.

Finally, an important reason to model how peer groups form is to be able to perform simulations in order to see if and how new peer groups begin after a redistribution program. Marmaros and Sacerdote (2006) use their poisson model coefficients to simulate different housing programs at Dartmouth by moving students around, which they find to be very small because the geographic effect on peer group formation is only relevant over small distances, so it is difficult to get students close to a large number of other students conditional on the dormitory structure of Dartmouth. For this chapter and the data in general, schools are generally large, so any simulated redistribution program would not run into proximity constraints. However, Marmaros and Sacerdote do find that changing the entire composition of the Dartmouth class does change peer groups along many different dimensions. The results of this chapter can be used in the construction of a structural model to attempt similar redistribution and composition changing simulations. For example, school redistribution can be simulated by redistributing disadvantaged students into a high achieving school district. The peer groups formed after the redistribution can then be analyzed.

1.3 Data-The National Longitudinal Study of Adolescent Health (Add Health)

Add Health is a nationally representative study that explores the causes of health related behaviors of adolescents in grades 7 through 12 and their outcomes into young adulthood (Udry 2003). It seeks to examine how social contexts (families, friends, peers, schools, neighborhoods, and communities) influence adolescents' health and risk behaviors.

The study was initiated in 1994 with an In-School survey administered to a nationally representative sample of about 90,000 students in grades 7 through 12 in 132 schools. These schools were selected to ensure that region, population density, size, type, and ethnicity were representative of the national population. It was administered in one day in a 45-60 minute class period. Questions asked include those about social and demographic characteristics of respondents, self-reported grades, education and occupation of parents, household structure, risk behaviors, expectations for the future, self-esteem, health status, friendships, and school-year extracurricular activities.

The friendship nominations will be exploited in this chapter. Each individual in the surveys could nominate up to five male friends and five female friends, and they were asked to rank friends in order of preference. The friends can either be from the individual's current school, a sister school, or from neither the current nor sister schools. About 15% of the friendship nominations in the In-School survey are not in the current or sister school of the individual, while about 8% of the nominations are from the individual's current school but not on the school's roster.

All students who completed the In-School survey plus those who did not complete a survey but were listed on a school roster were eligible for selection into the core In-Home sample. It is a sample of adolescents in grades 7 through 12 in the US during the 1994-95 school year. The survey is clustered around schools (although students took the survey at home). Students in each school were stratified by grade and gender. About 17 students were randomly chosen from each stratum so that a total of approximately 200 adolescents were selected from each of the 80 pairs of schools. A total core sample of about 12,000 adolescents was interviewed. A special oversample of well educated blacks, Chinese, Cuban, Puerto Rican, disabled, and some sibling pairs brought the total number of those who completed the In-Home survey in Wave I to about 20,000. In addition to the questions asked in the In-School survey, the In-Home survey asks additional questions on health status, health-facility utilization, nutrition, decision-making processes, family composition and dynamics, educational aspirations and expectations, employment experience, romantic experiences, substance use, and criminal activities. An aptitude test called the Add Health Picture Vocabulary Test (AHPVT) was administered before the survey. The AHPVT employs a series of images and words that describe these images in order to measure aptitude. The student must pick the word that best describes the picture. For example, a picture of a furry dog could be displayed to the user, with the words "furry," "greasy," "slimy," or "smooth."[3] Wave II of the In-Home Survey took place in 1996, and about 15,000 of the original respondents were retained. Wave III took place in 2001-2002, at which time the original respondents were asked about their current life situations. Topics included were current friends, drug usage, romantic relationships, and other activities. Additionally, they completed another AHPVT test.

1.4 Descriptive Statistics, Measures and Sorting Patterns

1.4.1 Descriptive Statistics

Sorting into friendships can potentially occur along many different dimensions. Race is often assumed to be a primary dimension, but others such as school performance and attitudes can affect friendship formation as well. Tables 1.1 and 1.2 describe the observations of the sample along race, academic, and attitudinal lines for the entire population using the In-School survey. Although sampling weights, stratification rules, and clustering procedures are provided in the data set, they were not used to construct these summary statistics. Therefore, these sample statistics should only be interpreted as accurate for the sample only, and not the general population.

Asians have the highest Grade Point Average (GPA) in English, math, sciences, and social studies, followed closely by whites.[4] Blacks and Hispanics subsequently have lower self-reported GPA's. It is interesting to note that blacks have a higher confidence in their

[3]The AHPVT standardized score ranges from 13-146.

[4]GPA's are standard, where A=4.0 and F=0.0.

Table 1.1: Sample Statistics, In-School Survey

Race	Percentage
White Only	52.03%
Black Only	15.42%
Asian Only	4.63%
Hispanic Only	17.22%
American Indian Only	0.93%
Other Only	1.66%
Mixed (more than one race)	5.92%
Nothing (no race specified)	2.20%
Number of Observations	89940

Table 1.2: Sample Statistics, In-School Survey, Part 2

		Total	White	Black	Asian	Hispanic
GPA	> 3.3	24.53%	29.32%	13.53%	39.09%	14.2%
	2.3-3.3	37.92%	37.6%	40.56%	38.82%	36.42%
	1.3-2.3	28.36%	25.1%	35.9%	18.08%	35.79%
	0.3-1.3	8.82%	7.65%	9.71%	3.8%	12.99%
	0-0.3	0.38%	0.33%	0.29%	0.21%	0.61%
Likely to	0-Unlikely	4.25%	3.72%	3.00%	2.27%	6.71%
Graduate College	1	1.29%	1.30%	0.95%	0.32%	1.65%
	2	5.64%	4.46%	6.88%	3.25%	9.54%
	3	1.23%	1.29%	0.94%	0.57%	1.49%
	4	8.89%	8.37%	8.48%	6.06%	11.65%
	5	2.25%	2.44%	1.45%	1.55%	2.18%
	6	14.76%	15.27%	13.26%	13.33%	15.29%
	7	10.43%	12.59%	6.20%	10.25%	7.38%
	8-Certainly	51.07%	50.56%	58.84%	62.41%	44.13%
Happy to	1-Strongly Agree	24.76%	25.54%	22.70%	23.54%	25.54%
Attend School	2	32.90%	33.96%	29.34%	37.03%	33.75%
	3	23.59%	22.87%	25.49%	25.47%	22.77%
	4	9.46%	9.16%	11.40%	7.82%	8.68%
	5-Strongly Disagree	9.29%	8.47%	10.99%	6.15%	9.26%

Table 1.3: Sample Statistics for the Representative Population, In-Home Survey

Race	Percentage
White Only	65.18%
Black Only	14.52%
Asian Only	0.57%
Hispanic Only	12.27%
American Indian Only	3.33%
Other Only	0.77%
Mixed (more than one race)	3.34%
Nothing (no race specified)	0.00%
Number of Observations	13568

future success in college, despite the grades. Asians are the most confident in their future collegiate success, while Hispanics have the least amount of confidence. The students' happiness at their school is reasonably even across races. The percentages of those in each race who responded positively to their school are close. However, there is a disparity between blacks and Asians who strongly disagree that they are happy at the school, with about 10% of blacks admitting that they are strongly unhappy as opposed to only about 6% of Asians having similar feelings.

Tables 1.3 and 1.4 describe the observations of only those in both wave 1 and wave 2 of the In-Home survey. The means reported in the table are population weighted to reflect sampling procedures (including the oversample). While most statistics seem to reflect the uncorrected values in the In-Home survey, there are a couple of exceptions.

The biggest exception is the severe undersampling of Asians, which even after correcting for the population weights, stratification, and clustering, does not reflect the uncorrected values for Asians in the In-School survey. Also, the descriptive statistics for Asians (not shown in Tables 1.3 and 1.4) do not reflect the values in the In-School survey. In fact, they are much lower achievement-wise. For these and other reasons, Asians are dropped from this analysis.[5]

1.4.2 Measures- Friends

The In-School and all three waves of the In-Home surveys ask for friendship nominations. There is no data for those friends who are not on the school roster, those friends who are from a sister school, or those friends who are not in the school. Therefore, these friends are excluded from the analysis. Each male and female friend must be from the same school as the respondent and must have been surveyed as well. Of the 44,811 males in the In-School survey, 31,535 have at least one male friend who is on the individual's current school roster, and of the 44,401 females in the In-School sample, 35,688 have at least one female friend who is listed. Of the 7,190 males in the In-Home survey who are listed in both wave 1 and wave 2, 4,254 have at least one male friend listed, and of the 7,546 females listed in wave 1 and wave 2, 4,529 have at least one female friend listed.

From the friendship nominations, a binary variable on whether an individual has a friend of a certain characteristic may be constructed. The analysis is limited to same-gender friends for simplicity and to avoid some confounding factors such as romantic relationships. If romantic and platonic relationships across races and achievement do not follow the same patterns, then platonic friendship results can be biased by including romantic partners among friends.[6] Concentrating on same-gender friendships eliminates most romantic relationship possibilities.[7]

The In-Home survey allows respondents' friendship nominations to include those who have taken the In-School survey, but have not taken the In-Home Survey. Since respondent identifiers are consistent throughout all the surveys and waves, it is possible to back out

[5]Arcidiacono and Nathan (2007) are working on describing and structurally modeling sorting and statistical discrimination patterns using the wave 1 In-School survey, in which Asians are used in the analysis.

[6]There are patterns in interracial romances that belie the general population (Foeman and Nance 1999).

[7]Another assumption is that male-female platonic friendships follow similar patterns to same-gender platonic friendships.

Table 1.4: Sample Statistics for the Representative Population, In-Home Survey, Part 2

		Total	White	Black	Asian	Hispanic
GPA-Wave 1	>3.3	22.95%	27.22%	11.30%	7.95%	12.88%
	2.3-3.3	41.01%	39.91%	45.07%	29.11%	40.19%
	1.3-2.3	28.53%	25.53%	36.73%	53.66%	36.97%
	0.3-1.3	7.08%	6.88%	6.35%	9.29%	9.45%
	0	0.45%	0.46%	0.55%	0.00%	0.50%
GPA-Wave 2	>3.3	21.60%	25.45%	9.56%	3.73%	12.17%
	2.3-3.3	37.82%	37.47%	41.29%	37.25%	36.75%
	1.3-2.3	30.77%	27.98%	38.48%	45.98%	36.70%
	0.3-1.3	9.38%	8.64%	10.15%	13.02%	14.17%
	0	0.44%	0.47%	0.52%	0.00%	0.21%
Mean AHPVT Standardized Score		100.77	104.62	91.42	98.38	92.71
Likely to Graduate College-Wave 1		51.19%	49.76%	58.83%	32.29%	46.08%
Likely to Graduate College-Wave 2		38.61%	40.90%	37.27%	14.29%	27.18%
Happy to Attend School-Wave 1		26.37%	27.04%	22.00%	18.23%	29.37%
Happy to Attend School-Wave 2		22.75%	23.79%	19.59%	20.05%	22.27%

relevant characteristics from the In-School survey that otherwise would not be available in just the In-Home survey.

1.4.3 Measures- Self-reported GPA vs. AHPVT as an Achievement Measure

Two potential measures for achievement include self-reported GPA and AHPVT score. There are advantages and disadvantages to using each measure. The advantage to using GPA is due to the number of people who report their grades in the In-School survey. As explained above, it is possible to back out the GPA of friends nominated in the In-Home survey, but not a respondent in the In-Home survey itself, from the In-School survey. The disadvantages to using GPA are that it can potentially be a noisy measure of achievement and that wave 2 GPA's for those friends nominated in the In-Home survey are unavailable.

AHPVT, on the other hand, has the advantage of generally being constant over time, since, like an IQ test, it attempts to appraise innate intelligence. The AHPVT is only taken by those in the In-Home survey in wave 1. However, it is plausible to suggest that an AHPVT score in wave 2 would be very similar to the AHPVT in wave 1. The big disadvantage to using the AHPVT is that the In-School survey cannot be used to back out the scores of friends nominated in the In-Home survey who were not respondents in the In-Home survey. Therefore, the number of friendship nominations who have AHPVT scores is significantly less.

As a result of these issues, this chapter uses GPA as the main measure for achievement. For those friendship nominees who did not take the In-Home survey in wave 2 (and thus have no GPA in wave 2), the GPA from wave 1 is carried over to wave 2.

1.4.4 Measures- Probabilities

In order to see if individuals in the sample are sorting across racial, academic, and attitudinal lines, it is important to compare the friendships that are actually formed with a random assignment of friendships in each school, where the friendship nominations in the sample originate. For example, group A may not have much interaction with group B due to the fact that they are not often in the same setting, so the probability that they form a random friendship is remote. However, if the probability that an individual from group A actually forms a friendship with an individual from group B is different from the probability of a random friendship conditional on the setting and characteristics, it may signal a sorting pattern. Sorting into a certain category is implied by a higher actual probability of friendship formation than the corresponding random probability of friendship formation, while sorting away from a certain category is implied by the opposite relationship.

- Actual $Pr(B|A)$ = Friendship probability from the data
- Random $Pr(B|A)$ = Indiscriminate matching probability within a school

Tables 1.5-1.8 list the actual and chance probabilities for respondents in the In-Home sample, divided on various lines, having listed same-gender friends along those same lines. The actual probabilities of friendship are calculated straight from the sample conditional on the individual's characteristics in the table. The random friendship probabilities are calculated by taking the mean of the relevant characteristic by school (since all friendships are contained within the school in the sample), conditional on an individual's characteristics in the table. The ratio of the actual probability to the random probability is reported. Any difference in the number of observations between the actual and random probabilities comes from individuals who are in the school population who do not list any valid friends.

In Tables 1.5 and 1.6, the actual probabilities that a respondent has a same-gender friend of the same race are higher for all races than the probability of randomly having a same-gender friend of the same race. Whites self-segregate less with regards to race than both blacks and Hispanics. In both waves, blacks are heavily sorting away from whites, due to the fact that the actual probability of a black student naming a white student as a same-gender friend is much lower than the probability of random interactions. Whites have a higher ratio of actual probability to random probability with regards to having Hispanic friends when compared to the ratio with respect to black friends. There is not a large difference between waves when comparing racial sorting. Academic and attitudinal

Table 1.5: Sorting Along Racial Lines (Wave I)

Respondent	Probability	White	Have Same Gender Friend Black	Hispanic
White	actual	92.77%	1.24%	4.93%
	random	80.65%	11.02%	7.90%
	ratio	1.15	0.11	0.62
Black	actual	5.80%	83.09%	10.54%
	random	28.89%	60.98%	9.73%
	ratio	0.20	1.36	1.08
Hispanic	actual	26.03%	4.42%	65.86%
	random	34.94%	16.11%	47.64%
	ratio	0.75	0.27	1.38
GPA>3.3	actual	83.23%	6.38%	6.34%
	random	73.46%	15.52%	10.45%
	ratio	1.13	0.41	0.61
Good College Prospects	actual	75.63%	12.65%	8.38%
	random	67.33%	20.70%	11.40%
	ratio	1.12	0.61	0.74
Happy at School	actual	77.94%	9.31%	9.77%
	random	69.55%	18.16%	11.75%
	ratio	1.12	0.51	0.83

sorting follow expected patterns, where higher achievers and those with good attitudes about college and their happiness in school sort into having white friends, and sort away from blacks and Hispanics.[8] As a quick test to see along what lines (racial, academic, and attitudinal) sorting dominates, similar probabilities were calculated using the In-School survey, but with the columns labeled as having a same-gender friend who has a high GPA, has good college prospects, and is happy at the school. Whites sort slightly into high academic achievement friendships, while blacks and Hispanics sort away from these types of friendships. Also, whites seem to sort into friendships with positive attitudes about school, while blacks and Hispanics sort away from friendships with positive attitudes about school. The ratios between the actual and random probabilities of interaction in those cases are much closer to 1 than when friends are separated by race. In fact, the sorting patterns observed in these descriptive tables imply that racial sorting may drive academic and attitudinal sorting. Overall, the patterns of sorting illustrated imply that racial sorting occurs more than academic sorting, which in turn is more prevalent than attitudinal sorting.

Tables 1.7 and 1.8 take the strongest patterns of sorting (racial and academic lines) and make each item both racially and academically dependent in wave 1 and wave 2 of the In-Home survey, respectively. Same-gender friends are categorized by their race and their achievement based on GPA together. Once again, there is sorting within an individual's own category. For whites and blacks, high achievers tend to self-segregate themselves at a higher rate than low black and white achievers in both wave 1 and wave 2. In fact, the degree of self segregation that occurs for high achieving blacks is almost triple that of the degree of self-segregation for low-achieving blacks, as measured by comparing ratios between the groups. This result is similar to one in Bayer, Fang and MacMillan (2005), where highly educated blacks self-segregate more than blacks who are less educated. Perhaps the small sample size of high-achieving blacks who actually have legitimate friends for analysis (61 in wave 1, 46 in wave 2) may be skewing results. A check was instituted using the In-

[8] In this particular context, an individual with a GPA of 3.3 (B+) or above is considered a "high" achiever, while an individual with GPA below 3.3 is considered a "low" achiever.

Table 1.6: Sorting Along Racial Lines (Wave II)

Respondent	Probability		Have Same Gender Friend	
		White	Black	Hispanic
White	actual	92.40%	1.17%	4.87%
	random	80.42%	11.19%	7.94%
	ratio	1.15	0.16	0.61
Black	actual	4.66%	85.15%	9.68%
	random	28.60%	61.33%	9.64%
	ratio	0.16	1.39	1.00
Hispanic	actual	29.72%	4.33%	62.26%
	random	34.67%	16.52%	47.47%
	ratio	0.86	0.26	1.31
GPA>3.3	actual	83.07%	5.97%	7.27%
	random	74.35%	14.57%	10.63%
	ratio	1.12	0.41	0.68
Good College Prospects	actual	74.63%	13.73%	8.47%
	random	67.95%	19.73%	11.84%
	ratio	1.10	0.70	0.72
Happy at School	actual	76.05%	11.56%	9.75%
	random	66.95%	18.95%	13.57%
	ratio	1.14	0.61	0.72

School survey, where there are no sample size problems.[9] The degree of self-segregation among high-achieving blacks was confirmed to be higher than the degree of self-segregation among low-achieving blacks in that sample as well. High-achieving Hispanics who have actual legitimate friendships also have a sample size problem here, as there are instances where there are no friends who are high-achieving Hispanics. However, when comparing similar statistics with the In-School survey, it is true that it is very unlikely, for example, for a low-achieving white to have a high-achieving Hispanic as a friend. The same is true regarding other cells that are empty in the In-Home waves (i.e. high-achieving whites and low-achieving Hispanics are very unlikely to actually have a high-achieving black friend). Therefore, the sample size problem does not affect how actual and random probabilities are compared. Finally, when looking across waves, it seems that high-achieving whites are integrating more with other racial/academic groups (actual probability of 52.69% of self-segregation in wave 1 compared to 44.26% in wave 2). The opposite is happening for high-achieving blacks (actual probability of 28.68% of self-segregation in wave 1 compared to 44.22% in wave 2), although this result may be driven by small sample sizes. There is not much change in low achievers across all racial groups with regards to self-segregation over time. Overall, descriptive statistics lend credence to the fact that homophily actually does exist and that it transcends both race and academics. Now, formal models of both homophily and statistical discrimination can be developed.

[9]Results in this chapter are compared to the In-School survey that is not corrected for survey design issues.

Table 1.7: Sorting Along Racial and Academic Lines (Wave I)

Respondent	Probability	Have Same Gender Friend					
		White >3.3	White <=3.3	Black >3.3	Black <=3.3	Hispanic >3.3	Hispanic <=3.3
White >3.3	actual (N=473)	52.69%	42.82%	0.06%	0.68%	1.78%	0.95%
	random (N=1805)	26.31%	54.34%	1.43%	9.74%	1.40%	6.37%
	ratio	2.00	0.79	0.04	0.07	1.27	0.15
White <=3.3	actual (N=1170)	23.16%	72.26%	0.00%	0.62%	0.19%	3.45%
	random (N=5043)	20.28%	60.82%	1.27%	9.45%	1.10%	6.66%
	ratio	1.14	1.19	0.00	0.07	0.17	0.52
Black >3.3	actual (N=61)	0.84%	2.26%	28.68%	63.68%	0.00%	3.36%
	random (N=347)	8.27%	20.42%	12.52%	49.03%	1.39%	8.10%
	ratio	0.10	0.11	2.29	1.30	0.00	0.42
Black <=3.3	actual (N=358)	1.05%	6.56%	7.51%	75.77%	0.09%	8.57%
	random (N=2213)	7.75%	21.14%	7.49%	53.72%	8.14%	1.41%
	ratio	0.14	0.31	1.00	1.41	0.01	6.09
Hispanic >3.3	actual (N=48)	13.00%	14.37%	0.00%	10.60%	15.45%	34.03%
	random (N=264)	12.71%	25.58%	2.56%	15.83%	7.35%	33.96%
	ratio	1.02	0.56	0.00	0.67	2.10	1.00
Hispanic <=3.3	actual (N=428)	6.54%	12.15%	0.15%	6.93%	8.01%	61.82%
	random (N=1899)	9.99%	25.45%	2.23%	13.97%	6.17%	40.94%
	ratio	0.65	0.48	0.07	0.50	1.30	1.51

Table 1.8: Sorting Along Racial and Academic Lines (Wave II)

Respondent	Probability	Have Same Gender Friend					
		White >3.3	White <=3.3	Black >3.3	Black <=3.3	Hispanic >3.3	Hispanic <=3.3
White >3.3	actual (N=355)	44.26%	48.48%	0.14%	0.62%	0.79%	3.71%
	random (N=1530)	24.51%	57.11%	1.08%	9.11%	1.31%	6.54%
	ratio	1.81	0.85	0.13	0.07	0.60	0.57
White <=3.3	actual (N=1040)	21.62%	71.81%	0.00%	1.64%	0.00%	4.20%
	random (N=4789)	18.82%	61.96%	1.15%	10.06%	1.00%	6.58%
	ratio	1.15	1.16	0.00	0.16	0.00	0.64
Black >3.3	actual (N=46)	0.00%	8.19%	44.22%	46.09%	0.00%	1.50%
	random (N=280)	6.83%	19.55%	9.99%	53.30%	1.40%	8.61%
	ratio	0.10	0.00	4.43	4.43	0.00	0.17
Black <=3.3	actual (N=357)	0.13%	2.37%	7.33%	82.19%	0.24%	7.11%
	random (N=2125)	6.72%	22.01%	7.07%	54.32%	1.12%	8.36%
	ratio	0.02	0.11	1.04	1.51	0.21	0.85
Hispanic >3.3	actual (N=29)	1.67%	20.02%	0.00%	2.62%	9.90%	65.79%
	random (N=207)	12.06%	28.56%	1.73%	16.60%	7.47%	33.85%
	ratio	0.14	0.70	0.00	0.16	1.32	1.94
Hispanic <=3.3	actual (N=328)	0.00%	28.11%	0.41%	5.49%	4.97%	57.66%
	random (N=1749)	8.98%	26.09%	2.43%	14.38%	5.57%	41.16%
	ratio	0.00	1.08	0.17	0.38	0.89	1.40

1.5 Homophily

1.5.1 Model

According to tables 1.2 and 1.4, it is apparent that whites on average have stronger GPA's than their black and Hispanic counterparts. Tables 1.5-1.8 show that there are heavy interactions within race and within academic achievement. Assume that there is a large influx of one race into another school, for example. Will the subsequent changing of the distribution of academic achievement within the school, in addition the the change in racial composition, cause changes in friendship formation within a school?

Let the probability of an individual in a certain school that has a same-gender friend of a certain racial group ($Prob(Y_{ijk})$) be represented by the following equation.[10]

$$Prob(Y_{ijk}) = \alpha_0 + X_i\alpha_1 + SHARE_{jk}\alpha_2 + (SHARE_{jk})^2\alpha_3 + \epsilon_{ijk} \tag{1.1}$$

- i = Individual respondent, j = School, k = Relevant racial group of friends
- X_i = Personal characteristics
- $SHARE_{jk}$ = Share of the relevant racial group in a school

There could potentially be nonlinear (most likely decreasing) returns to having more of a particular group at a school, which can be measured by including the squared term on the group shares at a school variable. So, it is expected that α_2 should be positive, while α_3 should be negative in order to confirm the decreasing returns to scale hypothesis. In order to test whether academics matters when sorting into friendship groups, the following addition can be made to the above equation.

$$Prob(Y_{ijk}) = \alpha_0 + X_i\alpha_1 + SHARE_{jk}\alpha_2 + (SHARE_{jk})^2\alpha_3 + (GRADE_i - \overline{GRADE}_j)\alpha_4 + \epsilon_{ijk} \tag{1.2}$$

- $GRADE_i$ = Academic metric for an individual i
- $\overline{GRADE}_j$ = Average of the academic metric at the school j, where i attends

[10]This model is outlined in Arcidiacono, Khan, and Vigdor (2008)

$(GRADE_i - \overline{GRADE}_j)$ is a measure of academic achievement for the individual relative to the school. It can measure if the student is an above or below average student relative to the school that the individual is enrolled. If it is the case that this measure does affect friendship formation, α_4 should be significantly different from zero. If α_4 is positive, than higher-achieving students are sorting into the friendship group of the race in question (k). If α_4 is negative, then higher-achieving students are sorting away from the friendship group of the race in question. If α_4 is zero, then homophily along academic achievement lines is insignificant in facilitating cross-race relationships.

1.5.2 Results

Tables 1.9 and 1.10 provide probit estimates of the above equation in waves 1 and 2, respectively. The dependent variable is an indicator of whether an individual in a racial group that is not the race in question ($-k$) has a same-gender friend of the race in question (k), and that race in question is either white, black, or Hispanic. X_i is represented in this case by gender, race, and attitudinal variables such as how the individual views his prospects for college and to what degree the individual is happy with experiences at school. The academic metric analyzed here is the individual's GPA. The coefficient on group shares (α_2) is positive and significant on all groups in both waves, which is expected. The coefficient on the square of group shares (α_3) is not significant for any groups across the two waves, so there is no evidence in these samples that there are decreasing returns to scale.[11] The individual characteristic (X_i) attitudinal variables are insignificant. In both waves, blacks have a negative and significant coefficient compared to other races on the probability of having a white friend. In wave 1, blacks have a positive, although insignificant, coefficient compared to other races on the probability of having a Hispanic friend. However, this coefficient is negative in wave 2. This lends some credence to the fact that blacks are self-segregating more in wave 2 than wave 1, and the difference can be weakly attributed to the switching of Hispanic friends to black friends. The coefficient on relative GPA also

[11] There is evidence of decreasing returns to scale in the In-School survey.

follows expected patterns regarding signs. It is positive and significant for those who have white friends in wave 1. Therefore, if a student is above-average relative to schoolmates academically, this student is more likely to have a same-gender friend who is white. The opposite effect is true when analyzing the coefficient on relative grades when the relevant friendship racial group is black or Hispanic. The coefficients are negative, meaning that if a student is above-average relative to schoolmates academically, then the student is less likely to have a same-gender friend who is black or Hispanic. Since the individuals in these regressions exclude the racial group of the friends in question, and through racial dummy variables which takes away any homophily effects of across races, homophily based on GPA can be isolated. So, on average, increasing the relative GPA of non-white students in a school has a positive effect on the probability of having a white friend, while increasing the relative GPA of non-black or non-Hispanic students in a school has a weakly negative effect on the probability of having a black or Hispanic friend respectively. Since whites in general have higher GPA's amongst these races, followed by Hispanics and blacks, it seems like homophily along GPA lines can facilitate cross-race friendships. All coefficients in wave 2 are not significant, but do have the expected signs. This result may be attributable to the GPA noise and carryover that is mentioned previously. A robustness check using a random effects probit with the "group" variable being the individual shows that coefficients on relative GPA follow the expected patterns. This method eliminates any factors that the individuals and schools may have that affect homophily over the two waves, such as any sort of academic tracking (most plausibly) and other institutions such as clubs (less plausibly).

These tables, along with the above descriptive tables with the actual and random probabilities, all suggest that similarities in characteristics associated with academic achievement (as well as attitudes to a lesser extent) seem to at least have a weak effect on friendships within and across races.

Table 1.9: Estimates on Having Friends from Various Groups Homophily (Wave I)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jk}$	1.605*	2.504***	2.071*
	(0.725)	(0.470)	(0.718)
$(SHARE_{jk})^2$	-0.554	-1.066	0.549
	(1.094)	(0.651)	(0.736)
$(GPA_i - \overline{GPA}_j)$	-0.0591	-0.0702	0.193*
	(0.0670)	(0.0446)	(0.0939)
Good College Prospects	-0.0120	-0.117	0.148
	(0.118)	(0.0800)	(0.112)
Happy at School	0.0691	-0.0135	0.0404
	(0.123)	(0.0804)	(0.112)
Male	0.115	0.00495	0.273*
	(0.108)	(0.0613)	(0.113)
Black		0.143	−1.084***
		(0.262)	(0.261)
Hispanic	−0.796*		-0.251
	(0.378)		(0.267)
White	−1.249**	-0.147	
	(0.388)	(0.254)	
Constant	−1.336**	−1.576***	-1.813
	(0.404)	(0.260)	(0.266)

Standard errors in parenthesis

* $p < .05$, ** $p < .01$, *** $p < .001$

† Race corresponds to column

†† All races except race in column are used in estimation

Table 1.10: Estimates on Having Friends from Various Groups Homophily (Wave II)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE_{jk}^{\dagger}$	1.703**	2.172***	2.487**
	(0.617)	(0.578)	(0.806)
$(SHARE_{jk})^2$	0.934	-0.702	-0.410
	(1.003)	(0.816)	(0.815)
$(GPA_i - \overline{GPA}_j)$	-0.108	-0.0393	0.00529
	(0.0719)	(0.0499)	(0.0811)
Good College Prospects	0.0771	-0.0499	0.244
	(0.145)	(0.0726)	(0.132)
Happy at School	0.112	-0.0368	0.235
	(0.115)	(0.0962)	(0.147)
Male	0.0973	0.170	-0.0917
	(0.141)	(0.0946)	(0.137)
Black		-0.112	−0.941**
		(0.181)	(0.357)
Hispanic	0.117		0.0482
	(0.360)		(0.342)
White	-0.293	-0.323	
	(0.353)	(0.164)	
Constant	−2.416***	−1.544***	−1.873***
	(0.356)	(0.196)	(0.395)

Standard errors in parenthesis

* $p < .05$, ** $p < .01$, *** $p < .001$

† Race corresponds to column

†† All races except race in column are used in estimation

1.6 Statistical Discrimination

1.6.1 Model

A test of statistical discrimination can be constructed as follows (Arcidiacono, Khan, and Vigdor 2008). Consider the share and share-squared variables in equation 1. In that equation, the share includes everyone in the group k (the race in question). These individuals in k can be split into those who have a better measure of achievement than individual i, those who have a similar measure of achievement to individual i, and those who have a worse measure of achievement than individual i.

$$SHARE_{jk}\alpha_2 = SHARE_{jkB}\alpha_{2B} + SHARE_{jkS}\alpha_{2S} + SHARE_{jkW}\alpha_{2W} \tag{1.3}$$

- $SHARE_{jkB}$ = Share of students in school j and group k who have a better academic achievement metric than individual i
- $SHARE_{jkS}$ = Share of students in school j and group k who have a similar academic achievement metric than individual i
- $SHARE_{jkW}$ = Share of students in school j and group k who have a worse academic achievement metric than individual i

Equation 1.3 is simply splitting the $SHARE_{jk}\alpha_2$ variable and coefficient into three tiers of academic achievement relative to the individual. The share is still relative to the entire population of the school, not just of the race in question. If the coefficients α_{2B}, α_{2S}, and α_{2W} are carried into the squared term as well, equation 1.1 becomes the following.

$$\begin{aligned} Prob(Y_{ijk}) = \alpha_0 + X_i\alpha_1 + SHARE_{jkB}\alpha_{2B} + SHARE_{jkS}\alpha_{2S} + SHARE_{jkW}\alpha_{2W} \quad (1.4) \\ +(SHARE_{jkB}\alpha_{2B} + SHARE_{jkS}\alpha_{2S} + SHARE_{jkW}\alpha_{2W})^2\alpha_3 + \epsilon_{ijk} \end{aligned}$$

The reason that the linear share coefficients enter into the squared term is to make sure that tiers with minimal first order effects (linear term) on the probability of having a friend in group k will also have minimal second order effects on the same probability. For example, if a certain tier does not have a large effect on the probability of having a friend in k, then an increase in the share of that tier should also be ensured not to have any effect on returns to scale, which is now purely measured by α_3.

If the tiering based on academic achievement is not important, then the coefficients on all three share variables should be the same. If the coefficient on the share of students in k who are better than the individual is higher than the coefficient on the share of students in k who are worse than the individual, and k has a measure of achievement that is lower than other races not in k, then the following is clear. Those individuals not in k are much more likely to have a friend in k if they are surrounded by high achieving members of k. In essence, individuals not in the group in question $(-k)$ happen to project the academic characteristics of k in their school (j) onto those students who could be potential friends. In this case, those individuals who are not in k are statistically discriminating on the basis of academic achievement against group k. Now, if the coefficient on the share of students in k who are worse than the individual is higher than the coefficient on the share of students in k who are better than the individual, and the measure of academic achievement for those not in k is lower than those in k, the opposite effect happens than mentioned above. However, once again, individuals not in the group in question $(-k)$ project characteristics of the $k's$ in their school (j) onto potential friends. This phenomenon also is an example of statistical discrimination against k by those not in k. Finally, the coefficient on the share of k that is similar in academic achievement to i can be used to measure the degree of homophily on academic achievement, since a projection of similar achievement to those not in k is placed on potential friends who happen to be in k. In summary, the estimation results of equation 4 can lead to the following.

- $\alpha_{2B} > \alpha_{2W}$, and $\overline{GRADE}_{-k} > \overline{GRADE}_k \Rightarrow$ Statistical Discrimination
- $\alpha_{2W} > \alpha_{2B}$, and $\overline{GRADE}_k > \overline{GRADE}_{-k} \Rightarrow$ Statistical Discrimination
- α_{2S} is a measure of homophily

1.6.2 Results

Tables 1.11 and 1.12 provide non-linear probit estimates for waves 1 and 2 of equation 4 and the marginal effects of a change in one standard deviation of the individual share variables of racial group k on the probability of having a same-gender friend in k, if the respondents are not in k.[12] Whites, blacks, and Hispanics are the racial groups that are analyzed. Achievement is once again measure by GPA. The tiers are defined as follows.

- Better: More than 0.5 GPA points above individual i
- Similar: Within 0.5 GPA points of individual i
- Worse: More than 0.5 GPA points below individual i

As implied in tables 1.2 and 1.4, blacks and Hispanics have lower GPA's in general than the population average, and whites have higher GPA's in general than the population average. Table 1.11 shows that there does exist statistical discrimination against blacks and Hispanics by non-blacks and non-Hispanics respectively in wave 1. A one standard deviation increase in the share of high-achieving blacks will result in an increase in the probability of having a black friend by 1.95% for non-blacks, while the corresponding probability increase that results from a one standard deviation increase in low-achieving blacks is 0.97%. With regards to having a Hispanic friend, the probability increases by 3.63% with a one standard deviation increase in the share of high-achieving Hispanics. The probability increases by 1.04% with a one standard deviation increase in the share of low-achieving Hispanics, but the estimate is insignificant. There is no evidence of statistical discrimination against whites, as the coefficients on the share of high-achieving whites and low-achieving whites are about the same, and the probabilities of having a white friend for non-whites change between 7% and 9% . In wave 1, there are decreasing returns to scale on the probabilities of having a Hispanic or white friend, but the coefficient on the share-squared coefficient is insignificant for blacks.

[12] A maximum likelihood estimation procedure outlined in Gould, Pitblado, and Sribney (2006) was used. The marginal effects reported are the average of all individual marginal effects.

Table 1.11: Estimates on Having Friends from Various Groups Statistical Discrimination (Wave I)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jkB}$	2.606*	5.382*	3.383*
	(1.001)	(1.995)	(1.795)
Marg. Effect (1 sd change)	1.95%	3.63%	9.45%
$SHARE^{\dagger}_{jkS}$	3.282*	4.744*	6.478*
	(1.071)	(1.088)	(0.909)
Marg. Effect (1 sd change)	2.46%	3.20%	15.89%
$SHARE^{\dagger}_{jkW}$	1.292*	1.534	3.208*
	(0.449)	(0.924)	(1.549)
Marg. Effect (1 sd change)	0.97%	1.04%	7.87%
$SHARE^2$	-0.0177	-0.216***	-0.113***
	(0.203)	(0.0367)	(0.0128)
Male	0.0376	0.000511	0.251*
	(0.108)	(0.0561)	(0.0925)
Black		-0.00848	-1.047***
		(0.0937)	(0.123)
Hispanic	-0.299*		-0.215*
	(0.121)		(0.100)
White	-0.789***	-0.311**	
	(0.138)	(0.103)	
Constant	-1.777***	-1.392***	-1.777***
	(0.140)	(0.113)	(0.179)

Standard errors in parenthesis

* $p < .05$, ** $p < .01$, *** $p < .001$

† Race corresponds to column

†† All races except race in column are used in estimation

Table 1.12 shows similar patterns exhibited in wave 2 as in wave 1, but magnitudes of statistical discrimination have lessened somewhat against blacks and Hispanics. The range in the probability of having a black friend for non-blacks goes from a 1.69% increase with a one standard deviation increase in high-achieving blacks to a 0.88% increase with a one standard deviation change in low-achieving blacks. The range in the probability of having a Hispanic friend for non-Hispanics goes from a 3.09% increase with a one standard deviation increase in high-achieving Hispanics to a 1.57% increase with a one standard deviation increase in low-achieving Hispanics. For whites, there is a slight shift towards being weakly statistically discriminated against by non-whites. The probability of having a white friend for non-whites increases by 9.61% with a one standard deviation increase in the share of low-achieving whites, while the probability increases by 8.37% with a one standard deviation increase in the share of high-achieving whites. The difference is slight. Homophily across both waves 1 and 2 seem to be prevalent, since the coefficients on shares that are similar to the GPA's of individuals are significant. However, the estimates of homophily here may be inflated due to the generally normal distribution of GPA's across the population.[13] To once again control for factors such as academic tracking and clubs, a robustness check using a random effects probit model with the "group" variable as the individual supports results that there exists statistical discrimination against blacks and Hispanics, but not against whites.[14]

1.7 Conclusion

This chapter has shown that homophily is very prevalent along racial lines and somewhat prevalent along academic lines. It has also shown that statistical discrimination along academic lines exists against blacks and Hispanics by non-blacks and non-Hispanics and has

[13] One way to correct for the inflation is to change the boundaries of better/similar/worse to a non-fixed number, such as deciles, instead. This option can be explored in future work.

[14] A model similar to equation 1.4 was estimated, except without the embedded coefficients in the squared term.

Table 1.12: Estimates on Having Friends from Various Groups Statistical Discrimination (Wave II)

	Non-Blacks with Black Same-Gender Friend††	Non-Hispanics with Hispanic Same-Gender Friend††	Non-Whites with White Same-Gender Friend††
$SHARE^{\dagger}_{jkB}$	3.214*	4.236*	3.961*
	(1.304)	(1.875)	(1.176)
Marg. Effect (1 sd change)	1.69%	3.09%	8.37%
$SHARE^{\dagger}_{jkS}$	6.215*	3.962*	6.390*
	(1.353)	(1.267)	(1.093)
Marg. Effect (1 sd change)	3.28%	2.89%	13.50%
$SHARE^{\dagger}_{jkW}$	1.676*	2.148*	4.549*
	(0.666)	(0.954)	(1.570)
Marg. Effect (1 sd change)	0.88%	1.57%	9.61%
$SHARE^2$	-0.0822	-0.141	-0.132^*
	(0.0496)	(0.109)	(0.0496)
$SHARE^2$	-0.0822	-0.141	-0.132***
	(0.0496)	(0.109)	(0.0128)
Male	-0.00590	0.152	-0.0114
	(0.127)	(0.0883)	(0.112)
Black		-0.0196	-1.018***
		(0.130)	(0.151)
Hispanic	-0.287		0.00307
	(0.161)		(0.132)
White	-0.653***	-0.207	
	(0.132)	(0.131)	
Constant	-2.033***	-1.626***	-1.734***
	(0.157)	(0.131)	(0.182)

Standard errors in parenthesis

* $p < .05$, ** $p < .01$, *** $p < .001$

† Race corresponds to column

†† All races except race in column are used in estimation

developed a measure of the magnitude of the statistical discrimination incurred. Finally, it has shown that the degree of statistical discrimination is also decreasing between wave 1 and wave 2, suggesting that the signal that is sent out by potential friends regarding academic achievement is becoming clearer, and each individual who chooses a friend stereotypes less as this signal becomes clearer.

The major policies that involve redistribution along racial lines are school redistricting and affirmative action. In principle, these policies assume a certain randomness in interracial contact based on the sheer number of students of a racial group in a certain institution. Therefore, any peer effect benefits that may be garnered from the interracial contact is often analyzed based on this often assumed randomness in peer group formation.

The results of this chapter show how non-randomness in peer group formation can be explained, which in turn can influence any peer effects from these groups. For example, take a policy which redistricts high-achieving minorities (in the case of this chapter, high-achieving black and Hispanic students) from poor school districts into better school districts with relatively few minorities.[15] The results above show that these minorities would experience statistical discrimination if the minorities who are already in the advantageous school district are low achievers. Therefore, the redistricted minority students would not integrate very well with the majority. My results, though, show that the signal of achievement put forth by the new minority students can become clearer after some time, and the degree of statistical discrimination based on academic achievement can decrease.

Future work would involve adding some structure to the above reduced form analysis. For example, individuals can get a random utility based on how "close" they are to potential friends based on racial/academic variables. The signal that is received by individuals from the potential friends is a mix of the true signal from the friends and some noise. Simulations can then be run based on various racial/academic assignment policies that assign friends to individuals, and the composition of friendship networks can then be analyzed.[16] Other

[15]See footnote 1 regarding the Metco program.

[16]A similar model is being thought about with regards to Arcidiacono and Nathan (2007), and something similar has been attempted in Arcidiacono, Khan, and Vigdor (2008).

future work involves, as mentioned previously, redefining tiers on a relative scale instead of an absolute scale to potentially remove bias when estimating homophily in equation 4.

Racial integration is both an end and a means to an end with regards to redistribution policies. This chapter analyzes how redistribution affects peer group composition, but the group composition's relation to the actual peer effects of the policy, such as future labor market outcomes or happiness due to the increase in diversity of these programs, is not analyzed. However, the composition of groups will certainly have an effect. This chapter shows that group composition forms in complex ways.

Chapter 2

The Effects of Racial and Extracurricular Friendship Diversity on Achievement

2.1 Introduction

The previous chapter explored certain mechanisms, such as sorting and statistical discrimination, that are used as tools in friendship formation. While friendship formation itself is a very interesting process to uncover, it would lose of lot of meaning if there were not some ramifications on how one's friends affect the choices, behavior, and achievement of one's self. This chapter will move towards the direction of estimating peer effects instead of discerning the mechanism of how certain friendships form.

This chapter attempts to answer if there is an effect of having different types of friends on achievement. For this analysis, an individual's achievement in GPA terms is estimated using an individual's friend's race or his racial friendship network heterogeneity and whether the individual's best same-gender friend is involved in similar extracurricular activities.

It is important to attempt an answer for the above problem due to many policy implications that may arise out of the issue. For example, school redistricting, which is an idea explored in chapter 1, usually has the effect of creating diversity (through channels such as income, race, and behavior) for both the child who is being redistricted as well as for the school and the new peers of the redistricted child. If students on a college campus form cliques that only associate with a homogenous group, such as Asians only having Asian friends or football players always only associating with each other, should more forceable integration schemes be put into action? Perhaps it is best to keep the segregation going due to some returns of being around a homogenous environment. Perhaps student athletes and regular students should be apart even more than they already are, or perhaps they should live in the same dormitory. An answer to the question on how diversity (in this case, specifically racial and extracurricular diversity) affects outcomes such as achievement would certainly help in assessing the policies just listed.

Just like the previous chapter, this chapter uses the National Longitudinal Study of Adolescent Health (Add Health) to take advantage of rich and real friendship networks in order to test various peer effects. The variables of major interest are GPA, whether someone has a best same-gender friend who is also of the same race, a heterogenous index measure to discover how diverse an individual's network really is, and indicators on whether best same-gender friends participated in one or more activity categories or specific clubs with the individual. Due to problems with endogeneity of the peer effects, ordinary least squares (OLS) is not the optimal estimation method. This chapter uses the "grade within a school" average on the basis of the endogenous variables in order to instrument for those variables, removes any bias caused by correlation with a disturbance term, and identifies the model. After going through all the scenarios, it is found that as care is taken to control for endogeneity, the more it seems like racial heterogeneity becomes important. Another way to look at it is that racial homogeneity could become less important, which is what mainly happens to blacks, Hispanics to a degree, and Asians. However, the most powerful result is that by two measures of friendship (best same-gender friend and friendship network), whites perform better in a more heterogeneous environment. With regards to clubs, it is found that there is very little effect of the heterogeneity regarding a best same-gender friend also participating in an activity type on an individual's GPA. However, when breaking down the friends into those who actually share places in individual sports or individual academic-scholastic clubs, it is found that there is not much of an effect with regards to heterogeneity in sporting activities, but there is an academic benefit regarding homogeneity in academic-scholastic clubs.

The chapter is organized as follows. Section 2.2 reviews the recent literature on the subject. Section 2.3 goes meticulously through models, which builds to the one used to correct for the most bias in this instance. Section 2.4 describes the Add Health In-School survey further, describes variables, and presents summary statistics. Section 2.5 presents the results of the various models shown in section 2.3. Section 2.6 concludes the chapter and sets up the future. Appendix A has some additional tables of network data and model specifications that were a little extraneous to the chapter, but of interest nonetheless. It also contains first-stage regression coefficients for the instrumental variable (IV) regressions in the main body of the chapter. The results section also explains basically what instruments are potentially strong and trustworthy, what instruments are weak and possibly problematic, and how to interpret the results with this information.

2.2 Recent Literature

Diversity is defined by Nehring and Puppe (2002) as "an aggregate of the pairwise dissimilarities between [a set of] elements. (p. 1155)" Alesina and La Ferrara (2005) explore it in the job market. The costs of diversity are the "conflict of preferences, racism, and prejudices, (p. 762)" while the benefits are "varieties in abilities, experiences, and cultures. (p. 762)" These reasons can explain various trends that will be reported in this chapter, such as perhaps why there might be a return to having a best friend of the same race. Many of the seminal papers on peer effects were mentioned in section 1.2 of chapter 1. The same efforts and problems (endogeneity) exist in this chapter as well.

A paper that analyzes policies on diversity and achievement very much like the ones outlined in section 2.1 is written by Angrist and Lang (2004). It analyzes Boston's Metropolitan Council for Educational Opportunity (Metco) program, which is a school redistricting program in the Boston metropolitan area. They studied the effect of the influx of poor performing minority students on affluent white students. They did not find an effect, with either ordinary least squares estimates or instrumental variables estimates. They did find that there was an effect on minority third graders, but due to the strange and localized nature of these effects, they concluded that there were few peer effects there. The story in this chapter is a little bit different, in that in Add Health, there is no burst of new people from far away suddenly coming into the school. However, the advantage that Add Health does have is that peer groups are extremely well defined, as students have to identify their friends to a surveyor.

Hoxby (2000) uses a very similar method to the one used in this chapter to identify peer effects in a classroom. Hoxby uses small grade level gender and race variations over consecutive years in order to identify the actual peer effect on achievement level. This chapter uses “grade within a school” variation of race and activity participation.

Mihaly (2008) also uses “grade within a school” variation, but uses the dependent variable as an indicator on if an individual drinks or smokes. Mihaly regresses it on the average of the individual’s peers’ behavior on drinking and smoking. To fix the endogeneity problem between an individual’s behavior and peer behavior, the “grade within a school” averages of drinking and smoking are used as instruments. In fact, it can probably be argued that the instruments in that case are a bit more plausible than using a similar type of variation on the dependent variable which is in this chapter (GPA). The dependent, endogenous, and instrument variables in that paper are all based on the same behavior. However, results of the peer effects are somewhat similar across the two papers in that, when evolving from OLS to IV regressions, estimates sometimes swing across zero and become significant going the other way, which is what happens in this chapter to whites when dealing with racial heterogeneity/homgoeneity.

Regarding activity participation, Broh (2002) shows that sports activities promote social ties in scholastic experiences, and that participation in most extracurricular activities in general improves achievement in most areas. Broh does not explicitly control for the friendships in such activities like in this chapter, but the results achieved in Broh’s paper agree with this chapter on the benefits of extracurricular activities on their own. Braddock (1981) also agrees that sports are beneficial to academics.

2.3 Model

The model begins with a general educational production function of relevant independent variables available in the Add Health In-School Sample on GPA.

$$GPA_i = \alpha_0 + X_i\alpha_1 + R_i\alpha_2 + A_i\alpha_3 + F_{im}\alpha_4 + \epsilon_i \tag{2.1}$$

- i = Individual respondent, m = Relevant trait (race or activity)
- X_i = Personal characteristics
- R_i = Race indicator of the individual
- A_i = Extracurricular activities indicator of the individual
- F_{im} = Indicator on having a friend with certain characteristics
- ϵ_i = Error term

The dependent variable is the GPA of an individual student. X_i and R_i are generally characteristics that can be considered purely exogenous.[1] For the purposes of this analysis, club participation (A_i) will also be treated as an exogenous variable. This model can be estimated by OLS. The model above, though, is not without its issues, because it is implausible to believe that F_{im} is identified. For example, perhaps the effect that F_{im} imposes is actually due to peer GPA, which is not in the model. Therefore, it is necessary to adjust the model to the specification listed below.

$$GPA_i = \alpha_0 + X_i\alpha_1 + R_i\alpha_2 + A_i\alpha_3 + F_{im}\alpha_4 + I_j + I_k + \epsilon_i \tag{2.2}$$

- i = Individual respondent, j = School, k = Grade m = Relevant trait (race or activity)
- I_j = School fixed effects
- I_k = Grade fixed effects

Equation 2.1 has a problem in that there are many unobservables that can potentially bias estimation. At the school level, variables such as the quality of teachers, facilities in the

[1] R_i is separate from X_i due to expositional purposes, since racial peer effects are a large part of this chapter.

school, and wealth of the neighborhood can certainly contribute to GPA. The grade level unobservables can include specific cohort effects such as within-grade teacher quality and admissions criteria changes. Equation 2.2 incorporates fixed effects at both the school and the grade level in order to control for these effects at these levels. However, there is still an endogeneity problem with F_{im} (friendship variable). The correlation between having a friend of a certain trait may be correlated with unobservables, such as an individual's friend's GPA. For example, if an individual's friend's GPA is correlated with both the individual's GPA and the actual friendship characteristic in question, the estimate on having a friend with the trait in question can be biased. Specifically, in equation 2.2, if trait m is correlated with an unobservable, which also affects the dependent variable GPA, it is very possible that α_3 is either overestimated or underestimated due to the unobservable.

An attempt to mitigate the endogeneity of F_{im} can be made. A variable needs to be found that is only correlated to GPA through F_{im}, which can then serve as an instrument to correct a potential endogeneity problem. A two-stage least squares (2SLS) IV procedure can be instituted using the following specification.

$$F_{im} = \beta_0 + X_i\beta_1 + R_i\beta_2 + A_i\beta_3 + \overline{F_{jkm}}\beta_4 + I_j + I_k + \delta_i \tag{2.3}$$

$$GPA_i = \alpha_0 + X_i\alpha_1 + R_i\alpha_2 + A_i\alpha_3 + \hat{F}_{im}\alpha_4 + I_j + I_k + \epsilon_i \tag{2.4}$$

- $\overline{F_{jkm}}$ = School-grade average of having trait m for individual i
- $\hat{F}_{im}$ = Linear predictor of F_{im} from equation 2.3

Equation 2.3 is the first stage of the 2SLS specification. Instruments include all of the independent variables in equation 2.2, in addition to the exclusion restriction which is the average of friendship characteristic m within individual $i's$ grade k in individual $i's$ school j. The implicit assumption here is that the excluded variable has to be exogenous and provides enough variation to be a strong instrument. Basically, within grade k of school j, the average of trait m should only affect an individual $i's$ GPA through individual $i's$ friendship variable F_{im}.

2.4 Data

2.4.1 Data Description

The data source that is explored in this chapter is the In-School portion of the National Longitudinal Study of Adolescent Health (Udry 2003).[2] In this chapter, the sampling scheme of the In-School survey is taken into account in order to weight the sample up to a population level survey. The sampling design is as follows.[3] 132 schools were preselected with an unequal probability in order to keep costs down and manageable. Therefore, these schools are the primary stage units of the survey, and thus a clustering adjustment can be made to standard errors in order to correct for non-random sampling. Generally, clustering will tend to increase any standard errors in regressions and standard deviations in means. However, the sample is also stratified by region of the country, and the sampling weights offered reflect stratification ex-post.

The variables used in this analysis can be divided into categories according to the model in section 2.3. The dependent variable, GPA, is a measure of scholastic achievement. It is calculated the same way as in the previous chapter; it is the average of self-reported grades given in English, mathematics, science, and social studies, where an "A" is given a 4 and an "F" is given a 1. The variables that constitute X_i are as follows.

- Male: Indicates the individual's gender
- Coll. Mom: Indicates if the individual's mother has graduated from college or achieved a greater education[4]
- U.S.A.: Indicates if the individual was born domestically
- Live Dad: Indicates if the individual lives with a biological father[5]
- Health: Indicates if the individual reports having excellent health
- Unsafe: Indicates if the individual feels safe in the individual's neighborhood[6]

[2] There is an overview of the survey introduced in section 1.3.

[3] Chantala and Tabor (1999) provide a very nice guide on the Add Health sampling scheme.

[4] This type of variable is often used as a proxy for family income when it is not available.

[5] It is a potential indicator on family quality.

[6] This type of variable also can be used as a proxy for family income when it is not available.

- Care: Indicates if the individual feels that the mother cares for the individual[7]
- TV: Indicates if the individual watches one hour or less of television in a day[8]

The variables under R_i constitute races. The races are divided into the categories of white, black, Asian, Hispanic, and other.[9] The reason why R_i is not included in X_i is that races are a key friendship characteristic analyzed in this chapter. It is obviously exogenous, though. The variables under A_i constitute being a member of a type of extracurricular activity. For the purposes of this analysis, the extracurricular activities are divided into two categories as defined in table 2.1.

In this chapter, individual club participation is taken as given. Also, note that the types of club participation are not mutually exclusive. Individuals may partake in any number of the activities in each of the categories. However, the following categories are constructed which are mutually exclusive.

- Sport: Indicates if an individual participates in ONLY sporting activities
- Acad-Schol.: Indicates if an individual participates in ONLY academic or scholastic activities
- Mixedclub: Indicates if an individual participates in both sporting and academic-scholastic activities
- Noclub: Indicates if an individual participates in no activities

Add Health has branded individuals who have listed as being a participant in more than 10 activities as unreliable. For example, there are individuals who claim they participate in all the sports offered in the list of activities. Therefore, these individuals are dropped from the analysis.

The relevant peer effect variables F_{im} are presumed to be endogenous. The four variables dealing with race that are used in this analysis consist of the following.

---

[7]This variable can capture a families unobserved environment.

[8]This variable can capture a student's unobserved motivation and study habits.

[9]Anybody who marked down having any sort of Hispanic lineage is considered a Hispanic for this analysis. The other race category includes Native Americans, mixed races, "other" races, and not marking a race.

Table 2.1: List of Clubs by Category

Academic-Scholastic	Sports
French Club	Baseball/Softball
German Club	Basketball
Latin Club	Field Hockey
Book Club	Football
Computer Club	Ice Hockey
Debate Team	Soccer
Drama Club	Swimming
Future Farmers of America	Tennis
History Club	Track
Math Club	Volleyball
Science Club	Wrestling
Band	Other Sport
Cheerleading/Dance	
Chorus/Choir	
Orchestra	
Newspaper	
Honor Society	
Student Council	
Yearbook	
Other Club	

- Same Race Fr.: Indicates whether an individual's best same-gender friend shares the individual's race[10]
- S. Hetero.: A heterogeneity index on an individual's send-network
- R. Hetero.: A heterogeneity index on an individual's receive-network
- S/R. Hetero.: A heterogeneity index on an individual's send/receive-network

The friendship variables are defined and asked in basically the same way as in the previous chapter. An individual can name up to five male and five female friends. It is encouraged by the surveyors that the best friend is included first. An individual's send-network is comprised of the set of individuals who an individual nominates as a friend as well as the individual. The number of ties in the network includes the number of friends who the individual nominates as well as any ties between those individuals, along with any friendship reciprocation by nominees back to the individual. For example, if individual i nominates friends p and q, these two ties are in $i's$ send-network. In addition, if p nominates q as a friend, then that tie is also in the $i's$ send-network. An individual's receive-network is analogous. It is comprised of the set of individuals who nominate i as a friend. The ties in the network include the ties to i along with ties among those who nominate i as a friend, as well as if i reciprocates a nomination. An individual's send/receive-network is the union of the send-network and the receive-network.

The following measure of racial heterogeneity is constructed by Add Health.[11]

$$Heterogeneity_{iR} = 1 - \left[\sum_{m=1}^{n} \left(\frac{R_m}{en} \right)^2 \right] \tag{2.5}$$

- R_m = Number of individuals with same race (m) as i
- en = Total unique members of the network that have valid data on R
- n = Total number of traits of R in the network

[10] The reason for treating friends as such is similar to the reason brought up in the last chapter. There are some inter-racial matters that confound the effects of having a cross-gender friend (Foeman and Nance 1999).

[11] The structure of the formula is the same for all the three types of networks. What changes are the numbers in the equation, based on the counts of the three networks.

An individual who has a very homogenous racial network will have a score approaching 0. An individual with a very heterogenous network will have a score that approaches 1.[12] The send-network is an accurate description of the individual's network as the individual perceives it. The receive-network is an accurate description of the individual's network as the people surrounding the individual perceive it. This chapter reports the send-network regressions in the results section, but receive-network and send/receive network regressions are reported in Appendix A.

The peer effect variables pertaining to extracurricular activities are also listed below.

- Same Act. Fr.: Indicates if an individual's best same-gender friend participates in the same activity category (Sports, Academic-Scholastic)
- Common Club Fr.: Indicates if an individual's best same-gender friend participates in at least one club WITHIN the same activity category

The distinction between the two variables lies in the fineness of the level of similarity. For example, a football player and an ice hockey player would share a similar activity (and would have the variable "Same Act. Fr." coded with a one), but would not share a common club (and would have the variable "Common Club Fr." coded as a zero). Perhaps effects would be stronger if the individuals are involved in common clubs, since they would be spending more time with each other than if they just shared an overall activity category. The results seem to back this hypothesis.

As outlined in section 3, the peer effect variables just discussed are endogenous. The relevant instruments used here are the "grade within school" average of similar races and activity type participation to the individual. For example, the instrument for having a best friend of the same race as well as the homogeneity index for a white individual is the average of whites in the individual's school, by grade. The instrument for having a best friend of the same activity category as well as the indicator on having a best friend in a shared specific club for an individual only involved in sports is the average of students who only participate in sports in the individual's school, by grade. The assumption here is that the

[12]Notice here that a network can be heterogenous to individual i even if all the others in the network are homogenous.

true effect on GPA is through the friend level variables, and not the "grade within school" variables. Sports participation across grades is actually quite variable. There is a clear trend that sports participation declines as an individual goes through high school. There is less variation with the other instruments across grades (race averages and potentially academic-scholastic club averages), so there is a danger of the instruments being suboptimal. Tables A.13 and A.14 in the appendix report the instrument coefficients and R^2 values for the first stages of the IV regressions reported in section 2.5.

2.4.2 Summary Statistics

Tables 2.2-2.4 display summary statistics for the relevant variables in this analysis. All means and standard deviations are corrected for probability weighting, stratification across regions, and clustering on schools, which brings these figures to the population level as opposed to the sample level. Observations are kept only if they have a value for the relevant variable, are in between 9th and 12th grade, and have valid population weights. The number N is the number of survey takers with relevant data in the appropriate cells. It is worth noting that there are about three times more whites than blacks or Hispanics, and that there are between three and four times more blacks and Hispanics than Asians in this survey.

Tables 2.2 and 2.3 cover the exogenous variables and how they vary along the four racial and two activity categories. Whites and Asians have the higher GPA's relative to blacks and Hispanics. Also, blacks have the highest participation rate in sporting activities alone, while Asians lead the way in participation when it comes to academic and scholastic activities alone. Only about 75% of Hispanics were born in the United States, while less than 50% of Asians were born in the United States. Asians are way ahead when it comes to mothers having at least a college degree, while blacks are well behind when it comes to biological fathers living at home.

Table 2.4 brings to light something encountered in the last chapter. That is, own-race friendships are clearly dominant over any friendships across races. With regard to

Table 2.2: Summary Statistics: GPA, Race and Activity Variables

	All	White	Black	Hisp.	Asian	Sport	Acad-Schol.
GPA	2.586	2.664	2.406	2.366	2.944	2.460	2.678
	(0.0294)	(0.0319)	(0.0482)	(0.0561)	(0.0396)	(0.0315)	(0.0287)
N	48672	28030	6476	7572	2324	10372	12303
White	0.581					0.579	0.602
	(0.0343)					(0.0324)	(0.0407)
N	59326					12610	14386
Black	0.158					0.170	0.157
	(0.0270)					(0.0264)	(0.0342)
N	59326					12610	14386
Hisp.	0.135					0.134	0.114
	(0.0205)					(0.0176)	(0.0215)
N	59326					12610	14386
Asian	0.0349					0.0244	0.0389
	(0.006279)					(0.00444)	(0.00725)
N	59326					12610	14386
Sport	0.213	0.212	0.230	0.212	0.149		
	(0.0675)	(0.00852)	(0.0143)	(0.0117)	(0.0104)		
N	59326	32205	8525	10278	2824		
Acad-Schol.	0.238	0.247	0.237	0.202	0.266		
	(0.0790)	(0.00891)	(0.0206)	(0.0115)	(0.0249)		
N	59326	32205	8525	10278	2824		

Standard deviations in parenthesis

Table 2.3: Summary Statistics: Other Exogenous Variables

	All	White	Black	Hisp.	Asian	Sport	Acad-Schol.
U.S.A.	0.923	0.981	0.968	0.732	0.481	0.933	0.930
	(0.0115)	(0.00228)	(0.00753)	(0.0415)	(0.0494)	(0.0104)	(0.0112)
N	57710	31526	8292	9881	2758	12264	14081
Coll. Mom	0.304	0.309	0.299	0.238	0.542	0.292	0.300
	(0.0147)	(0.0173)	(0.0295)	(0.0124)	(0.0345)	(0.0145)	(0.0164)
N	57029	31633	8004	9643	2730	12092	14059
Live Dad	0.762	0.833	0.541	0.710	0.839	0.752	0.770
	(0.0125)	(0.00731)	(0.0159)	(0.0135)	(0.0187)	(0.0138)	(0.0128)
N	57765	31873	8148	9821	2762s	12277	14187
Health	0.286	0.273	0.328	0.296	0.301	0.359	0.214
	(0.00566)	(0.00665)	(0.00918)	(0.0100)	(0.0148)	(0.00844)	(0.00706)
N	57710	31708	8043	9468	2741	12348	14232
Unsafe	0.0856	0.0743	0.0848	0.114	0.104	0.0766	0.0897
	(0.00315)	(0.00286)	(0.00778)	(0.00557)	(0.0111)	(0.00486)	(0.00502)
N	53773	30568	7350	8439	2582	11503	13582
Care	0.862	0.885	0.841	0.824	0.841	0.853	0.881
	(0.00403)	(0.00377)	(0.00740)	(0.0101)	(0.0199)	(0.00650)	(0.00418)
N	57425	31736	8092	9749	2743	12187	14110
TV	0.249	0.290	0.119	0.214	0.298	0.224	0.244
	(0.0122)	(0.0127)	(0.00851)	(0.00868)	(0.0496)	(0.0112)	(0.0110)
N	57244	31715	8045	9519	2747	12369	14257

Standard deviations in parenthesis

the heterogeneity measures, all of them show that Hispanics are the most heterogenous, followed by Asians and blacks, and then whites. Also, more than 50% of individuals who are only involved in sporting activities have their best same-gender friend involved in one or more of those activities as well.

2.5 Results

The tables and discussion in this section are arranged in the following order. Racial effects for both variables of interest (best same-gender friend and send-network racial diversity index) are presented first. Activity effects are then presented next. Within each category, OLS with fixed school and grade effects (equation 2.2), and IV estimation using 2SLS (equations 2.3 and 2.4) are presented. The column labeled "All" includes all non-missing observations between 9th and 12th grade. All other columns are conditional on the column header. In all regressions, whites are the omitted racial independent variable, while those individuals not participating in any activities comprise the omitted activity group. Therefore, all racial variable coefficients are relative to whites, while all activity category variable coefficients are relative to non-participants.

Tables 2.5 and 2.6 analyze how GPA is affected by an individual's best same-gender friend. After introducing OLS with fixed effects in table 2.5, the coefficient for whites is positive and weakly significant. Hispanics have a coefficient that is weakly significant on the negative side.

However, after instrumenting for the friend peer effect variables in table 2.6, whites have a highly significant negative coefficient on having a best same-gender friend who is also white. Also, even though they are insignificant, all friend coefficients for other races have negative point estimates. So, according to this specification, friendship diversification seems to increase GPA. With regard to the other explanatory variables, being involved in a club of some sort increases GPA, except for Asians playing sports. Also, females are certainly performing better than males through all races. The other exogenous variables

Table 2.4: Summary Statistics: Endogenous Variables

	All	White	Black	Hisp.	Asian	Sport	Acad-Schol.
Same Race Fr.	0.694	0.817	0.718	0.471	0.478	0.676	0.719
	(0.0118)	(0.00967)	(0.0260)	(0.0523)	(0.0513)	(0.0133)	(0.0134)
N	43772	25212	6019	6715	2050	9094	11566
White Fr.	0.596	0.817	0.0518	0.249	0.257	0.596	0.604
	(0.0330)	(0.00967)	(0.0107)	(0.0363)	(0.0456)	(0.0311)	(0.0410)
N	43772	25212	6019	6715	2050	9094	11566
Black Fr.	0.139	0.0141	0.718	0.0941	0.0290	0.134	0.150
	(0.0255)	(0.00196)	(0.0260)	(0.0182)	(0.00722)	(0.0212)	(0.0344)
N	43772	25212	6019	6715	2050	9094	11566
Hisp. Fr.	0.107	0.0478	0.0776	0.471	0.0949	0.115	0.0968
	(0.0154)	(0.00432)	(0.00720)	(0.0523)	(0.0140)	(0.0154)	(0.0164)
N	43772	25212	6019	6715	2050	9094	11566
Asian Fr.	0.305	0.0130	0.00580	0.0290	0.478	0.0248	0.0318
	(0.00638)	(0.00212)	(0.00143)	(0.006220	(0.0513)	(0.00522)	(0.00684)
N	43772	25212	6019	6715	2050	9094	11566
S. Hetero.	0.215	0.153	0.212	0.431	0.341	0.223	0.205
	(0.0112)	(0.00878)	(0.0188)	(0.0117)	(0.0282)	(0.0120)	(0.0120)
N	40285	23495	5639	5864	1895	8298	10778
R. Hetero.	0.210	0.143	0.212	0.425	0.340	0.212	0.202
	(0.0109)	(0.00808)	(0.0174)	(0.0127)	(0.0264)	(0.0110)	(0.0119)
N	44263	24948	6602	6805	2078	9255	11343
S/R. Hetero.	0.240	0.174	0.244	0.450	0.369	0.244	0.231
	(0.0122)	(0.00958)	(0.0198)	(0.0120)	(0.0274)	(0.0125)	(0.0132)
N	47074	26237	7071	7377	2271	9880	12000
Same Act. Fr.	0.467	0.470	0.461	0.494	0.408	0.529	0.322
	(0.00641)	(0.00663)	(0.0144)	(0.0142)	(0.0227)	(0.0148)	(0.00928)
N	41405	24187	5618	6134	1944	8593	11021
Common Club Fr.	0.419	0.442	0.375	0.374	0.438	0.380	0.425
	(0.00566)	(0.00594)	(0.00959)	(0.0100)	(0.0217)	(0.0107)	(0.0105)
N	43772	25212	6019	6715	2050	9094	11566
Sport Fr.	0.197	0.201	0.193	0.191	0.155	0.380	0.0771
	(0.00786)	(0.00973)	(0.0177)	(0.00942)	(0.0116)	(0.0106)	(0.00478)
N	43772	25212	6019	6715	2050	9094	11566
Acad-Schol. Fr.	0.235	0.237	0.250	0.222	0.251	0.0854	0.425
	(0.00881)	(0.103)	(0.240)	(0.0129)	(0.0259)	(0.00515)	(0.0105)
N	43772	25212	6019	6715	2050	9094	11566

Standard deviations in parenthesis

Table 2.5: Shared Race of Best Friend on GPA-OLS

	All	White	Black	Hisp.	Asian
Same Race Fr.	0.0263**	0.0369**	-0.00361	-0.0610*	0.0352
	(0.0150)	(0.0216)	(0.0365)	(0.0383)	(0.0600)
Black	-0.177***				
	(0.0281)				
Hisp.	-0.173***				
	(0.0269)				
Asian	0.186***				
	(0.0375)				
Male	-0.0947***	-0.0909***	-0.153***	-0.0654***	-0.137***
	(0.0205)	(0.0259)	(0.0325)	(0.0309)	(0.0416)
Sport	0.207***	0.248***	0.102***	0.121***	-0.0168
	(0.0192)	(0.0211)	(0.0425)	(0.0370)	(0.0949)
Acad-Schol.	0.391***	0.419***	0.274***	0.332***	0.339***
	(0.0210)	(0.0280)	(0.0436)	(0.0291)	(0.0740)
U.S.A.	-0.100***	-0.0788*	-0.0627	-0.0903***	-0.146***
	(0.0234)	(0.0492)	(0.0772)	(0.0397)	(0.0520)
Coll. Mom	0.163***	0.191***	0.0985***	0.0836***	0.0985***
	(0.0144)	(0.0168)	(0.0420)	(0.0341)	(0.0391)
Live Dad	0.118***	0.140***	0.131***	0.111***	0.190***
	(0.0139)	(0.0202)	(0.0285)	(0.0293)	(0.0681)
Health	0.139***	0.157***	0.0764***	0.141***	0.111***
	(0.0127)	(0.0165)	(0.0348)	(0.0348)	(0.0419)
Unsafe	-0.119***	-0.144***	-0.122***	-0.0851*	0.0591
	(0.0191)	(0.0245)	(0.0330)	(0.0513)	(0.0714)
Care	0.175***	0.163***	0.123***	0.173***	0.219***
	(0.0146)	(0.0218)	(0.0343)	(0.0496)	(0.0728)
TV	0.0992***	0.114***	0.0333	0.0193	0.0483
	(0.0153)	(0.0168)	(0.0538)	(0.0415)	(0.0361)
N	33175	20463	4136	4315	1594

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.6: Shared Race of Best Friend on GPA-IV

	All	White	Black	Hisp.	Asian
Same Race Fr.	-0.0835	-1.156***	-0.809	-0.339	-1.796
	(0.0796)	(0.456)	(1.295)	(0.497)	(1.807)
Black	-0.190***				
	(0.0292)				
Hisp.	-0.220***				
	(0.0453)				
Asian	0.145***				
	(0.0470)				
Male	-0.100***	-0.150***	-0.228**	-0.0734***	-0.223***
	(0.0206)	(0.0340)	(0.119)	(0.0329)	(0.0872)
Sport	0.209***	0.283***	0.110***	0.118***	-0.00482
	(0.0191)	(0.0313)	(0.0522)	(0.0372)	(0.105)
Acad-Schol.	0.392***	0.448***	0.272***	0.317***	0.428***
	(0.0209)	(0.0318)	(0.0470)	(0.0415)	(0.144)
U.S.A.	-0.108***	-0.0177	-0.0668	-0.124*	-0.319**
	(0.0250)	(0.0663)	(0.0821)	(0.0790)	(0.182)
Coll. Mom	0.163***	0.205***	0.101**	0.0786***	0.0534
	(0.0146)	(0.0199)	(0.0535)	(0.0367)	(0.0868)
Live Dad	0.118***	0.150***	0.131***	0.107***	0.159*
	(0.0141)	(0.0258)	(0.0287)	(0.0308)	(0.105)
Health	0.140***	0.172***	0.0720**	0.141***	0.186*
	(0.0124)	(0.0187)	(0.0369)	(0.0334)	(0.117)
Unsafe	-0.124***	-0.194***	-0.147***	-0.0977**	0.0295
	(0.0200)	(0.0368)	(0.0634)	(0.0556)	(0.107)
Care	0.176***	0.171***	0.119***	0.174***	0.154*
	(0.0145)	(0.0262)	(0.0374)	(0.0472)	(0.0949)
TV	0.0988***	0.116***	0.0257	0.0147	0.0121
	(0.0151)	(0.0169)	(0.0677)	(0.0450)	(0.0888)
N	33175	20463	4136	4315	1594

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

generally behave as expected for this regression and subsequent regressions reported below, except for being born in the U.S.A., which is generally either significant or has a very slight negative effect on GPA.

A similar pattern and direction hold true when the endogenous variable is changed to the heterogeneity index of an individual's send-network in tables 2.7 and 2.8. Whites go from having an insignificant index coefficient using OLS to a significantly positive index coefficient (meaning heterogeneity will increase GPA). The heterogeneity index for whites in this case has a standard deviation of about 0.153.[13] Since the coefficient for whites is slightly larger than 1.5, a one standard deviation change upwards in the heterogeneity index would yield a GPA increase of slightly above 0.2 points. For blacks and Asians, although coefficients are not significant, there is a shift in point estimates to the positive, which is in favor of heterogeneity over homogeneity with regards to GPA. Hispanics, however, go from being very weakly significant on the positive side with the OLS estimates to having no significance with the IV specification. Once again, Asians who play sports are the only club participants who have a negative (but insignificant) coefficient. Tables A.1-A.4 in the appendix list regressions with different friendship network variables. One set of tables has estimates using the receive-network, and the other set of tables has estimates using the send/receive-network.

Table 2.9 lists the OLS regression on having a best same-gender friend in the same activity category on GPA. The coefficient for those who only participate in academic or scholastic activities is weakly positively significant, meaning that homogeneity in friendship on activity lines mildly increases GPA. For the population at large (which includes those individuals who are in both types of activities and those individuals who are not in any activities), having a similar best same-gender friend on activity lines affects GPA positively. After instrumenting in table 2.10 with the "grade within school" average of similar club participation, however, all coefficients on having a best same-gender friend in the same activity category are insignificant.

[13] This number is calculated by taking the sample and keeping those who are between 9th and 12th grade as well as having non-missing values for the heterogeneity index.

Table 2.7: Send-Network Friendship Heterogeneity on GPA-OLS

	All	White	Black	Hisp.	Asian
S. Hetero.	0.0215	-0.0138	0.0611	0.135*	-0.0317
	(0.0397)	(0.0495)	(0.0750)	(0.0901)	(0.129)
Black	-0.172***				
	(0.0285)				
Hisp.	-0.178***				
	(0.0251)				
Asian	0.206***				
	(0.0360)				
Male	-0.0996***	-0.0943***	-0.138***	-0.0628**	-0.149***
	(0.0215)	(0.0266)	(0.0337)	(0.0359)	(0.0384)
Sport	0.220***	0.258***	0.112***	0.127***	0.00377
	(0.0194)	(0.0219)	(0.0445)	(0.0435)	(0.0917)
Acad-Schol.	0.396***	0.424***	0.276***	0.325***	0.348***
	(0.0220)	(0.0294)	(0.0424)	(0.0314)	(0.0740)
U.S.A.	-0.0907***	-0.0850*	-0.0491	-0.0797**	-0.117***
	(0.0241)	(0.0519)	(0.0839)	(0.0426)	(0.0486)
Coll. Mom	0.164***	0.190***	0.0971***	0.0846***	0.0881***
	(0.0147)	(0.0172)	(0.0445)	(0.0290)	(0.0420)
Live Dad	0.113***	0.123***	0.135***	0.119***	0.195***
	(0.0144)	(0.0214)	(0.0305)	(0.0330)	(0.0716)
Health	0.133***	0.156***	0.0665**	0.131***	0.102***
	(0.0134)	(0.0171)	(0.0358)	(0.0393)	(0.0401)
Unsafe	-0.114***	-0.136***	-0.132***	-0.0327	0.0859
	(0.0205)	(0.0257)	(0.0354)	(0.0581)	(0.0741)
Care	0.171***	0.163***	0.113***	0.160***	0.203***
	(0.0156)	(0.0228)	(0.0345)	(0.0531)	(0.0580)
TV	0.0988***	0.114***	0.0336	0.0398	0.0420
	(0.0159)	(0.0173)	(0.0574)	(0.0453)	(0.0415)
N	30589	19028	3908	3809	1464

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.8: Send-Network Friendship Heterogeneity on GPA-IV

	All	White	Black	Hisp.	Asian
S. Hetero.	0.149	1.574**	1.074	-2.158	1.871
	(0.174)	(0.800)	(1.823)	(12.37)	(2.763)
Black	-0.179***				
	(0.0307)				
Hisp.	-0.208***				
	(0.0459)				
Asian	0.189***				
	(0.0415)				
Male	-0.100***	-0.0906***	-0.175***	-0.0259	-0.190***
	(0.0214)	(0.0251)	(0.0744)	(0.212)	(0.0588)
Sport	0.220***	0.254***	0.109***	0.137***	-0.0549
	(0.0192)	(0.0236)	(0.0448)	(0.0612)	(0.104)
Acad-Schol.	0.395***	0.423***	0.253***	0.376	0.346***
	(0.0218)	(0.0294)	(0.0517)	(0.274)	(0.0674)
U.S.A.	-0.0930***	-0.0745	-0.0365	-0.0277	-0.233
	(0.0245)	(0.0543)	(0.0929)	(0.300)	(0.166)
Coll. Mom	0.165***	0.205***	0.0932**	0.0626	0.0505
	(0.0150)	(0.0198)	(0.0546)	(0.135)	(0.0774)
Live Dad	0.113***	0.121***	0.138***	0.128**	0.193***
	(0.0146)	(0.0267)	(0.0372)	(0.0642)	(0.0905)
Health	0.134***	0.158***	0.0651**	0.116	0.116***
	(0.0133)	(0.0173)	(0.0365)	(0.0921)	(0.0403)
Unsafe	-0.116***	-0.164***	-0.148***	0.00634	0.0928
	(0.0208)	(0.0286)	(0.0524)	(0.227)	(0.0824)
Care	0.172***	0.171***	0.110***	0.176**	0.122
	(0.0156)	(0.0244)	(0.0365)	(0.0887)	(0.116)
TV	0.0995***	0.122***	0.0353	0.0298	0.0382
	(0.0161)	(0.0185)	(0.0634)	(0.0635)	(0.0458)
N	30589	19028	3908	3809	1464

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.9: Shared Activity Category of Best Friend on GPA-OLS

	All	Sport	Acad-Schol
Same Act. Fr.	0.0436***	0.0361	0.0395**
	(0.0113)	(0.0273)	(0.0217)
Sport	0.206***		
	(0.0192)		
Acad-Schol.	0.388***		
	(0.0212)		
Male	-0.0948***	-0.105***	-0.113***
	(0.0204)	(0.0400)	(0.0297)
Black	-0.178***	-0.181***	-0.185***
	(0.0283)	(0.0480)	(0.0442)
Hisp.	-0.183***	-0.186***	-0.142***
	(0.0256)	(0.0375)	(0.0379)
Asian	0.176***	-0.0194	0.311***
	(0.0382)	(0.0701)	(0.0606)
U.S.A.	-0.104***	-0.0825	-0.109***
	(0.0234)	(0.0568)	(0.0326)
Coll. Mom	0.162***	0.138***	0.202***
	(0.0144)	(0.0235)	(0.0229)
Live Dad	0.118***	0.0897***	0.103***
	(0.0139)	(0.0339)	(0.0213)
Health	0.138***	0.152***	0.145***
	(0.0125)	(0.0287)	(0.0319)
Unsafe	-0.119***	-0.183***	-0.112***
	(0.0191)	(0.0472)	(0.0381)
Care	0.174***	0.179***	0.177***
	(0.0146)	(0.0386)	(0.0318)
TV	0.0987***	0.00450	0.135***
	(0.0154)	(0.0259)	(0.0265)
N	33175	6746	9231

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.10: Shared Activity Category of Best Friend on GPA-IV

	All	Sport	Acad-Schol
Same Act. Fr.	0.153	0.201	3.240
	(0.122)	(0.649)	(3.358)
Sport	0.203***		
	(0.0199)		
Acad-Schol.	0.286***		
	(0.0235)		
Male	-0.0918***	-0.148	0.525
	(0.0220)	(0.166)	(0.693)
Black	-0.167***	-0.195***	0.116
	(0.0259)	(0.0712)	(0.333)
Hisp.	-0.171***	-0.183***	0.0905
	(0.0293)	(0.0387)	(0.212)
Asian	0.144***	-0.0176	0.197
	(0.0377)	(0.0719)	(0.200)
U.S.A.	-0.108***	-0.0928	-0.0984
	(0.0245)	(0.0724)	(0.111)
Coll. Mom	0.151***	0.138***	0.160***
	(0.0141)	(0.0238)	(0.0602)
Live Dad	0.119***	0.0956***	0.0981**
	(0.0150)	(0.0433)	(0.0572)
Health	0.134***	0.148***	0.154***
	(0.0118)	(0.0287)	(0.0628)
Unsafe	-0.125***	-0.179***	-0.121
	(0.0194)	(0.0502)	(0.0881)
Care	0.166***	0.177***	0.113
	(0.0172)	(0.0411)	(0.131)
TV	0.0933***	0.00470	0.137***
	(0.0156)	(0.0261)	(0.0666)
N	30477	6746	6533

Standard errors in parentheses
Instrument = Grade within school average of activity
* $p < .15$, ** $p < .10$, *** $p < .05$

However, there is a change when looking at best friends who share participation in at least one common club within a specific activity category. Table 2.11 presents OLS results. Here, all coefficients on having a best same-gender friend in at least one common club are significant. In the "All" and the "Sport" category, more homogeneity with this type of friend leads to a slightly higher GPA, while more heterogeneity in the "Acad-Schol." category leads to a slightly higher GPA. After instrumenting with the same variables as in table 2.10, it actually turns out that individuals who only participate in academic or scholastic activities get a better GPA in homogenous friendships with their best friend (table 2.12). The coefficients in the other two columns turn out to be insignificant after instrumenting. As discussed in section 2.4, a possible reason for the differences in significance for the last two IV regressions (tables 2.10 and 2.12) can be attributed to a greater influence of having a best friend in the exact same club(s) versus just having a best friend in a similar activity category.

As an added test to see how prevalent friendship effects can be, the sample is divided into movers and non-movers.[14] Non-movers should have more significance on their peer effect variables, due to the fact that they have been around longer and potentially have established friendships. Tables 2.13 and 2.14 confirm this hypothesis, as the peer effect coefficients on non-movers are insignificant. The peer effect coefficients on movers generally follow the same significance patterns and directions as the full sample. When thinking of actual policies to change student behavior, it must be taken into account that they may have some lag to their effects, as tables 2.13 and 2.14 seem to indicate.

Tables A.13 and A.14 in the appendix have the instruments used in all the IV regressions, as well as all of the 1st stage coefficients on those instruments. Notice that some appear weak (especially on the Hispanic regressions), which could explain some insignificance in the IV regressions conditional on the relevant group. Also, there is a potential

[14]The "mover" variable is crudely constructed by first isolating individuals between 10th and 12th grade. If a 10th grader has been in the school for one year, an 11th grader in the school for one or two years, or a 12th grader in the school for one, two, or three years, then they are counted as movers. Note that there may be a measurement problem if schools are not traditionally separated between middle and high school in 9th grade.

Table 2.11: Shared Common Clubs of Best Friend on GPA-OLS

	All	Sport	Acad-Schol
Common Club Fr.	0.0689***	0.0902***	-0.0614***
	(0.0158)	(0.0274)	(0.0266)
Sport	0.240***		
	(0.0240)		
Acad-Schol.	0.440***		
	(0.0263)		
Male	-0.0838***	-0.0994***	-0.109***
	(0.0207)	(0.0390)	(0.0309)
Black	-0.179***	-0.189***	-0.183***
	(0.0287)	(0.0504)	(0.0450)
Hisp.	-0.192***	-0.185***	-0.176***
	(0.0251)	(0.0384)	(0.0375)
Asian	0.164***	0.00452	0.272***
	(0.0383)	(0.0702)	(0.0643)
U.S.A.	-0.105***	-0.0635	-0.129***
	(0.0244)	(0.0595)	(0.0330)
Coll. Mom	0.159***	0.134***	0.202***
	(0.0150)	(0.0241)	(0.0239)
Live Dad	0.117***	0.0786***	0.0971***
	(0.0138)	(0.0351)	(0.0223)
Health	0.133***	0.154***	0.133***
	(0.0130)	(0.0286)	(0.0325)
Unsafe	-0.125***	-0.185***	-0.103***
	(0.0187)	(0.0492)	(0.0380)
Care	0.177***	0.175***	0.184***
	(0.0152)	(0.0417)	(0.0335)
TV	0.0956***	0.00127	0.131***
	(0.0154)	(0.0275)	(0.0258)
N	31609	6415	8829

Standard errors in parentheses
* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.12: Shared Common Clubs of Best Friend on GPA-IV

	All	Sport	Acad-Schol
Common Club Fr.	0.407	0.188	0.753***
	(0.287)	(1.993)	(0.339)
Sport	0.400***		
	(0.136)		
Acad-Schol.	0.522***		
	(0.162)		
Male	-0.0717***	-0.109	-0.0173
	(0.0247)	(0.189)	(0.0631)
Black	-0.174***	-0.199	-0.129***
	(0.0260)	(0.195)	(0.0510)
Hisp.	-0.167***	-0.181***	-0.0335
	(0.0312)	(0.0783)	(0.0601)
Asian	0.142***	0.0168	0.230***
	(0.0367)	(0.257)	(0.0828)
U.S.A.	-0.106***	-0.0657	-0.157***
	(0.0252)	(0.0751)	(0.0385)
Coll. Mom	0.139***	0.131***	0.112***
	(0.0155)	(0.0493)	(0.0336)
Live Dad	0.115***	0.0777***	0.0764***
	(0.0149)	(0.0378)	(0.0311)
Health	0.117***	0.147	0.0861***
	(0.0169)	(0.139)	(0.0407)
Unsafe	-0.122***	-0.178	-0.104***
	(0.0175)	(0.150)	(0.0513)
Care	0.162***	0.171**	0.147***
	(0.0198)	(0.0874)	(0.0469)
TV	0.0872***	0.00218	0.0971***
	(0.0170)	(0.0290)	(0.0425)
N	29001	6415	6221

Standard errors in parentheses

Instrument = Grade within school average of activity

* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.13: Racial Friendship Regression Coefficients on Movers-IV

Movers	All	White	Black	Hisp.	Asian
Same Race Fr.	-0.144	0.959	0.222	-0.777	-5.150
	(0.175)	(0.963)	(1.999)	(4.300)	(12.02)
N	5477	2884	850	792	483
S. Hetero.	-0.149	-2.758	0.0329	-1.120	235.5
	(0.357)	(2.730)	(2.788)	(2.300)	(7820.7)
N	5081	2693	810	714	444
Non-Movers	**All**	**White**	**Black**	**Hisp.**	**Asian**
Same Race Fr.	-0.0939	-1.345***	-2.061	-0.298	-2.869
	(0.0779)	(0.523)	(1.486)	(0.369)	(2.939)
N	27698	17579	3286	3523	1111
S. Hetero.	0.244*	1.756***	2.259	5.826	2.801
	(0.165)	(0.834)	(1.557)	(39.07)	(3.312)
N	25508	16335	3098	3095	1020

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table 2.14: Activity Friendship Regression Coefficients on Movers-IV

Movers	All	Sport	Acad-Schol.
Same Act. Fr.	-0.0953	-0.525	3.055
	(0.212)	(1.250)	(3.974)
N	5083	1053	1098
Common Club Fr.	-0.105	-1.857	0.570
	(0.579)	(4.923)	(0.421)
N	4685	973	1013
Non-Movers	**All**	**Sport**	**Acad-Schol.**
Same Act. Fr.	0.129	0.290	4.849
	(0.114)	(0.651)	(5.928)
N	25394	5693	5435
Common Club Fr.	0.371	0.420	0.867***
	(0.298)	(1.858)	(0.399)
N	24316	5442	5208

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

weak instrument in each of the activity regressions, which include the sports regression on the best same-gender friend in a common club and the academic-scholastic regression on the best same-gender friend in a certain activity category. Not coincidentally, these IV estimates are insignificant. Therefore, paying attention to IV regressions with potentially strong instruments is much more beneficial for any sort of inference. In the appendix, tables A.5-A.12 remove an individual's club participation variables from all the regressions listed above. In some instances, the magnitudes rise, but all of the meaningful coefficients in the regressions on racial lines remain the same, so similar conclusions can be drawn.

2.6 Conclusion

Uncovering the achievement effects of different types of racial and extracurricular friendships is an idea that can be fraught with danger. The biggest danger is that the peer effect estimates may in fact be biased due to endogeneity problems. This chapter chronicles a model in steps from OLS with no fixed effects, to an OLS model with school fixed effects and grade fixed effects, and finally to a model that uses exogenous exclusion restrictions at the "grade within a school" level to instrument for endogenous peer variables. Using a rich data set in Add Health where peer networks can be constructed, it is possible to ascertain, with proper instrumenting, how much peer diversity along racial and extracurricular lines can actually help achievement.

The results suggest that, to some extent, OLS using fixed school and grade effects can be biased towards homogeneity in having friends, based on racial lines. Almost all racial peer variables were biased in the direction towards homogeneity, while instrumenting led to estimates that were insignificant on GPA or that even favored heterogeneity. In two IV regressions, whites could increase their GPA by a little less than 1.2 points by having same-gender friends of a different race. Somewhat less abruptly, an increase in the send-network heterogeneity of an individual white person by one standard deviation would increase GPA by just a little more then 0.2 points. Whites were clearly affected the most by the composition of their best same-gender friend's race. With regards to extracurricular activities, homogeneity yields higher GPA's among academic or scholastic activity participants, but it is inconclusive on how it affects those individuals who participate in sports activities. Results, though, may be tempered by weak instruments on some specifications.

Although there are certain strong results in this analysis, it would be premature to say that peer diversity would be the certain cause of higher achievement. However, future work to estimate causality better could involve the simulation of friendship assignments instead of taking the network as given in a reduced form specification, which is what this chapter presents. Nevertheless, the result of racial friendship heterogeneity increasing achievement (especially amongst whites) can potentially set a precedent of tolerance and understanding over isolation and segregation with regards to race.

Appendix A

Appendix: The Effects of Racial and Extracurricular Friendship Diversity on Achievement

Table A.1: Receive-Network Friendship Heterogeneity on GPA-OLS

	All	White	Black	Hisp.	Asian
R. Hetero.	0.0262	-0.0318	0.141**	0.167*	-0.101
	(0.0363)	(0.0455)	(0.0754)	(0.101)	(0.100)
Black	-0.172***				
	(0.0281)				
Hisp.	-0.171***				
	(0.0235)				
Asian	0.233***				
	(0.0360)				
Male	-0.0998***	-0.102***	-0.120***	-0.0773***	-0.117***
	(0.0208)	(0.0257)	(0.0319)	(0.0347)	(0.0362)
Sport	0.213***	0.255***	0.101***	0.155***	-0.0636
	(0.0188)	(0.0225)	(0.0447)	(0.0423)	(0.0782)
Acad-Schol.	0.389***	0.425***	0.309***	0.275***	0.257***
	(0.0200)	(0.0277)	(0.0423)	(0.0321)	(0.0569)
U.S.A.	-0.0790***	-0.0364	-0.0814	-0.0618**	-0.140***
	(0.0249)	(0.0499)	(0.0832)	(0.0313)	(0.0436)
Coll. Mom	0.160***	0.185***	0.0966***	0.0753***	0.104***
	(0.0142)	(0.0165)	(0.0375)	(0.0318)	(0.0470)
Live Dad	0.103***	0.118***	0.129***	0.0881***	0.176***
	(0.0160)	(0.0231)	(0.0291)	(0.0313)	(0.0587)
Health	0.135***	0.160***	0.0753***	0.0934***	0.111***
	(0.0124)	(0.0162)	(0.0344)	(0.0382)	(0.0363)
Unsafe	-0.122***	-0.144***	-0.112***	-0.0710	0.0925
	(0.0224)	(0.0250)	(0.0384)	(0.0540)	(0.0875)
Care	0.158***	0.159***	0.0909***	0.133***	0.190***
	(0.0159)	(0.0234)	(0.0272)	(0.0579)	(0.0739)
TV	0.0936***	0.105***	0.0500	0.0724**	0.0484
	(0.0145)	(0.0175)	(0.0535)	(0.0412)	(0.0405)
N	32574	19943	4329	4205	1556

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.2: Receive-Network Friendship Heterogeneity on GPA-IV

	All†	White	Black	Hisp.	Asian
R. Hetero.	0.176	1.782***	0.895	13.27	0.385
	.	(0.798)	(2.267)	(31.56)	(1.357)
Black	-0.183				
	(0)				
Hisp.	-0.208				
	(0)				
Asian	0.212				
	.				
Male	-0.0999	-0.0978***	-0.137***	-0.00313	-0.109***
	(0)	(0.0246)	(0.0559)	(0.187)	(0.0523)
Sport	0.213	0.251***	0.0941***	0.181	-0.0719
	.	(0.0250)	(0.0436)	(0.139)	(0.0713)
Acad-Schol.	0.388	0.409***	0.283***	0.236	0.266***
	.	(0.0265)	(0.0702)	(0.195)	(0.0608)
U.S.A.	-0.0829	-0.0275	-0.0716	-0.587	-0.172**
	(0)	(0.0495)	(0.0862)	(1.267)	(0.0943)
Coll. Mom	0.160	0.187***	0.0920***	0.0385	0.0898
	.	(0.0170)	(0.0425)	(0.164)	(0.0672)
Live Dad	0.104	0.119***	0.121***	0.475	0.167***
	.	(0.0307)	(0.0369)	(0.914)	(0.0684)
Health	0.135	0.164***	0.0696*	-0.0409	0.113***
	.	(0.0163)	(0.0445)	(0.325)	(0.0359)
Unsafe	-0.124	-0.173***	-0.115***	-0.0421	0.0971
	(0)	(0.0311)	(0.0428)	(0.161)	(0.0905)
Care	0.158	0.164***	0.0862***	0.145	0.186***
	.	(0.0254)	(0.0347)	(0.129)	(0.0722)
TV	0.0940	0.105***	0.0509	0.128	0.0418
	.	(0.0156)	(0.0568)	(0.161)	(0.0411)
N	32574	19943	4329	4205	1556

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

† Standard errors not calculable

Table A.3: Send/Receive-Network Friendship Heterogeneity on GPA-OLS

	All	White	Black	Hisp.	Asian
S/R. Hetero.	0.00414	-0.0668	0.153***	0.0582	-0.0710
	(0.0434)	(0.0565)	(0.0627)	(0.101)	(0.112)
Black	-0.170***				
	(0.0273)				
Hisp.	-0.170***				
	(0.0236)				
Asian	0.225***				
	(0.0343)				
Male	-0.103***	-0.104***	-0.132***	-0.0734***	-0.121***
	(0.0207)	(0.0260)	(0.0289)	(0.0348)	(0.0398)
Sport	0.215***	0.255***	0.114***	0.159***	-0.0481
	(0.0188)	(0.0218)	(0.0419)	(0.0380)	(0.0727)
Acad-Schol.	0.392***	0.425***	0.306***	0.297***	0.291***
	(0.0203)	(0.0280)	(0.0405)	(0.0291)	(0.0589)
U.S.A.	-0.0832***	-0.0573	-0.127**	-0.0541	-0.105***
	(0.0245)	(0.0501)	(0.0758)	(0.0380)	(0.0494)
Coll. Mom	0.162***	0.190***	0.0893***	0.0829***	0.0963***
	(0.0136)	(0.0160)	(0.0372)	(0.0282)	(0.0418)
Live Dad	0.108***	0.122***	0.131***	0.0953***	0.189***
	(0.0140)	(0.0199)	(0.0308)	(0.0294)	(0.0569)
Health	0.133***	0.159***	0.0687***	0.0994***	0.113***
	(0.0126)	(0.0169)	(0.0338)	(0.0358)	(0.0347)
Unsafe	-0.115***	-0.131***	-0.113***	-0.0685	0.0850
	(0.0218)	(0.0246)	(0.0357)	(0.0541)	(0.0824)
Care	0.158***	0.159***	0.0869***	0.129***	0.199***
	(0.0158)	(0.0226)	(0.0268)	(0.0554)	(0.0603)
TV	0.0939***	0.105***	0.0328	0.0693**	0.0714**
	(0.0141)	(0.0165)	(0.0512)	(0.0408)	(0.0355)
N	34509	20915	4626	4551	1702

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.4: Send/Receive-Network Friendship Heterogeneity on GPA-IV

	All†	White	Black	Hisp.	Asian
S/R. Hetero.	0.216	1.653***	0.865	8.893	0.471
	.	(0.794)	(1.781)	(20.78)	(1.320)
Black	-0.185				
	(0)				
Hisp.	-0.216				
	(0)				
Asian	0.201				
	.				
Male	-0.104	-0.0985***	-0.150***	-0.0584	-0.126***
	(0)	(0.0247)	(0.0507)	(0.0697)	(0.0383)
Sport	0.214	0.254***	0.107***	0.107	-0.0575
	.	(0.0243)	(0.0423)	(0.156)	(0.0658)
Acad-Schol.	0.390	0.415***	0.285***	0.190	0.297***
	.	(0.0278)	(0.0565)	(0.269)	(0.0625)
U.S.A.	-0.0869	-0.0497	-0.106	-0.290	-0.139*
	(0)	(0.0502)	(0.0867)	(0.555)	(0.0852)
Coll. Mom	0.163	0.200***	0.0821**	0.0975*	0.0810
	.	(0.0176)	(0.0452)	(0.0623)	(0.0584)
Live Dad	0.109	0.124***	0.131***	0.193	0.178***
	.	(0.0257)	(0.0328)	(0.240)	(0.0620)
Health	0.133	0.164***	0.0631*	0.108	0.111***
	.	(0.0170)	(0.0424)	(0.0801)	(0.0377)
Unsafe	-0.118	-0.163***	-0.125***	-0.137	0.0842
	(0)	(0.0265)	(0.0546)	(0.177)	(0.0808)
Care	0.158	0.164***	0.0847***	0.105	0.190***
	.	(0.0240)	(0.0287)	(0.120)	(0.0583)
TV	0.0951	0.112***	0.0331	0.182	0.0675**
	.	(0.0170)	(0.0533)	(0.304)	(0.0371)
N	34509	20915	4626	4551	1702

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

† Standard errors not calculable

Table A.5: Shared Race of Best Friend on GPA-OLS (NO CLUBS)

	All	White	Black	Hisp.	Asian
Same Race Fr.	0.0335***	0.0560***	-0.0105	-0.0845***	0.0429
	(0.0156)	(0.0238)	(0.0356)	(0.0396)	(0.0653)
Black	-0.170***				
	(0.0316)				
Hisp.	-0.181***				
	(0.0300)				
Asian	0.219***				
	(0.0396)				
Male	-0.155***	-0.153***	-0.205***	-0.118***	-0.204***
	(0.0186)	(0.0232)	(0.0283)	(0.0288)	(0.0386)
U.S.A.	-0.0936***	-0.0809*	-0.0675	-0.0848***	-0.132***
	(0.0249)	(0.0506)	(0.0789)	(0.0414)	(0.0527)
Coll. Mom	0.201***	0.234***	0.126***	0.0992***	0.113***
	(0.0159)	(0.0184)	(0.0457)	(0.0338)	(0.0390)
Live Dad	0.138***	0.168***	0.146***	0.129***	0.215***
	(0.0149)	(0.0220)	(0.0275)	(0.0295)	(0.0717)
Health	0.175***	0.203***	0.0835***	0.158***	0.126***
	(0.0130)	(0.0162)	(0.0340)	(0.0333)	(0.0424)
Unsafe	-0.140***	-0.175***	-0.134***	-0.0793*	0.0553
	(0.0200)	(0.0274)	(0.0294)	(0.0503)	(0.0816)
Care	0.193***	0.192***	0.118***	0.176***	0.236***
	(0.0162)	(0.0239)	(0.0330)	(0.0509)	(0.0846)
TV	0.116***	0.131***	0.0408	0.0300	0.0763**
	(0.0164)	(0.0176)	(0.0557)	(0.0432)	(0.0411)
N	33175	20463	4136	4315	1594

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.6: Shared Race of Best Friend on GPA-IV (NO CLUBS)

	All	White	Black	Hisp.	Asian
Same Race Fr.	-0.0868	-1.343***	-0.777	-0.515	-1.807
	(0.0914)	(0.431)	(1.291)	(0.482)	(1.767)
Black	-0.184***				
	(0.0343)				
Hisp.	-0.232***				
	(0.0527)				
Asian	0.174***				
	(0.0513)				
Male	-0.161***	-0.223***	-0.273***	-0.125***	-0.306***
	(0.0190)	(0.0309)	(0.116)	(0.0290)	(0.108)
U.S.A.	-0.101***	-0.00889	-0.0714	-0.138**	-0.306**
	(0.0265)	(0.0714)	(0.0833)	(0.0799)	(0.179)
Coll. Mom	0.201***	0.256***	0.127***	0.0897***	0.0690
	(0.0162)	(0.0222)	(0.0556)	(0.0387)	(0.0854)
Live Dad	0.139***	0.182***	0.144***	0.121***	0.188**
	(0.0151)	(0.0292)	(0.0272)	(0.0318)	(0.107)
Health	0.176***	0.228***	0.0789***	0.157***	0.199**
	(0.0128)	(0.0196)	(0.0366)	(0.0318)	(0.115)
Unsafe	-0.145***	-0.238***	-0.158***	-0.0992**	0.0244
	(0.0210)	(0.0403)	(0.0633)	(0.0530)	(0.110)
Care	0.194***	0.203***	0.115***	0.177***	0.171**
	(0.0163)	(0.0311)	(0.0388)	(0.0489)	(0.0965)
TV	0.116***	0.135***	0.0330	0.0217	0.0409
	(0.0163)	(0.0185)	(0.0678)	(0.0483)	(0.0864)
N	33175	20463	4136	4315	1594

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.7: Send-Network Friendship Heterogeneity on GPA-OLS (NO CLUBS)

	All	White	Black	Hisp.	Asian
S. Hetero.	0.0253	-0.0197	0.0979	0.176**	-0.0250
	(0.0456)	(0.0596)	(0.0804)	(0.0910)	(0.146)
Black	-0.166***				
	(0.0319)				
Hisp.	-0.189***				
	(0.0281)				
Asian	0.235***				
	(0.0385)				
Male	-0.159***	-0.156***	-0.190***	-0.115***	-0.214***
	(0.0194)	(0.0237)	(0.0303)	(0.0343)	(0.0384)
U.S.A.	-0.0864***	-0.0858*	-0.0571	-0.0754**	-0.102**
	(0.0257)	(0.0535)	(0.0872)	(0.0438)	(0.0509)
Coll. Mom	0.202***	0.234***	0.124***	0.103***	0.102***
	(0.0166)	(0.0190)	(0.0486)	(0.0302)	(0.0419)
Live Dad	0.132***	0.151***	0.148***	0.135***	0.214***
	(0.0154)	(0.0234)	(0.0302)	(0.0332)	(0.0753)
Health	0.170***	0.203***	0.0762***	0.148***	0.120***
	(0.0137)	(0.0162)	(0.0345)	(0.0374)	(0.0413)
Unsafe	-0.138***	-0.170***	-0.146***	-0.0369	0.0883
	(0.0219)	(0.0290)	(0.0314)	(0.0606)	(0.0843)
Care	0.188***	0.190***	0.109***	0.162***	0.223***
	(0.0178)	(0.0254)	(0.0333)	(0.0553)	(0.0744)
TV	0.115***	0.128***	0.0427	0.0529	0.0840**
	(0.0169)	(0.0179)	(0.0593)	(0.0460)	(0.0463)
N	30589	19028	3908	3809	1464

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.8: Send-Network Friendship Heterogeneity on GPA-IV (NO CLUBS)

	All	White	Black	Hisp.	Asian
S. Hetero.	0.145	1.694***	1.033	-0.182	1.825
	(0.203)	(0.757)	(1.843)	(10.37)	(2.407)
Black	-0.173***				
	(0.0360)				
Hisp.	-0.216***				
	(0.0561)				
Asian	0.220***				
	(0.0460)				
Male	-0.160***	-0.153***	-0.218***	-0.111	-0.261***
	(0.0193)	(0.0236)	(0.0663)	(0.131)	(0.0639)
U.S.A.	-0.0886***	-0.0743	-0.0451	-0.0672	-0.217
	(0.0260)	(0.0555)	(0.0961)	(0.250)	(0.154)
Coll. Mom	0.203***	0.251***	0.119***	0.0997	0.0644
	(0.0170)	(0.0210)	(0.0578)	(0.0984)	(0.0734)
Live Dad	0.132***	0.149***	0.149***	0.137***	0.212***
	(0.0156)	(0.0294)	(0.0353)	(0.0591)	(0.0914)
Health	0.171***	0.205***	0.0748***	0.146**	0.126***
	(0.0136)	(0.0171)	(0.0357)	(0.0737)	(0.0453)
Unsafe	-0.140***	-0.201***	-0.160***	-0.0308	0.0950
	(0.0224)	(0.0310)	(0.0501)	(0.196)	(0.0879)
Care	0.188***	0.199***	0.107***	0.165**	0.144
	(0.0179)	(0.0281)	(0.0372)	(0.0850)	(0.0985)
TV	0.116***	0.137***	0.0435	0.0517	0.0781*
	(0.0171)	(0.0194)	(0.0645)	(0.0529)	(0.0503)
N	30589	19028	3908	3809	1464

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.9: Receive-Network Friendship Heterogeneity on GPA-OLS (NO CLUBS)

	All	White	Black	Hisp.	Asian
R. Hetero.	0.0407	-0.0236	0.197***	0.179**	-0.103
	(0.0402)	(0.0545)	(0.0750)	(0.0991)	(0.0943)
Black	-0.167***				
	(0.0302)				
Hisp.	-0.186***				
	(0.0258)				
Asian	0.257***				
	(0.0375)				
Male	-0.163***	-0.170***	-0.186***	-0.116***	-0.174***
	(0.0197)	(0.0236)	(0.0313)	(0.0328)	(0.0388)
U.S.A.	-0.0814***	-0.0452	-0.0966	-0.0668***	-0.127***
	(0.0250)	(0.0510)	(0.0851)	(0.0290)	(0.0466)
Coll. Mom	0.199***	0.231***	0.125***	0.0930***	0.109***
	(0.0164)	(0.0190)	(0.0424)	(0.0344)	(0.0457)
Live Dad	0.125***	0.148***	0.144***	0.104***	0.190***
	(0.0170)	(0.0249)	(0.0297)	(0.0317)	(0.0598)
Health	0.174***	0.208***	0.0845***	0.113***	0.123***
	(0.0129)	(0.0153)	(0.0324)	(0.0364)	(0.0380)
Unsafe	-0.145***	-0.174***	-0.126***	-0.0800	0.0883
	(0.0240)	(0.0288)	(0.0384)	(0.0557)	(0.102)
Care	0.175***	0.189***	0.0940***	0.129***	0.208***
	(0.0192)	(0.0267)	(0.0264)	(0.0611)	(0.0851)
TV	0.108***	0.117***	0.0539	0.0809**	0.0872**
	(0.0149)	(0.0179)	(0.0552)	(0.0408)	(0.0449)
N	32574	19943	4329	4205	1556

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.10: Receive-Network Friendship Heterogeneity on GPA-IV (NO CLUBS)

	All	White	Black	Hisp.	Asian
R. Hetero.	0.168	1.840***	0.863	11.75	0.458
	(0.212)	(0.787)	(2.303)	(29.31)	(1.179)
Black	-0.177***				
	(0.0370)				
Hisp.	-0.218***				
	(0.0578)				
Asian	0.238***				
	(0.0467)				
Male	-0.163***	-0.163***	-0.197***	-0.0335	-0.168***
	(0.0197)	(0.0242)	(0.0445)	(0.196)	(0.0509)
U.S.A.	-0.0847***	-0.0356	-0.0870	-0.528	-0.164**
	(0.0247)	(0.0500)	(0.0897)	(1.166)	(0.0920)
Coll. Mom	0.199***	0.232***	0.119***	0.0565	0.0914
	(0.0163)	(0.0197)	(0.0470)	(0.152)	(0.0626)
Live Dad	0.125***	0.148***	0.136***	0.443	0.181***
	(0.0172)	(0.0326)	(0.0412)	(0.843)	(0.0677)
Health	0.173***	0.213***	0.0790**	-0.0103	0.124***
	(0.0130)	(0.0163)	(0.0444)	(0.318)	(0.0401)
Unsafe	-0.147***	-0.204***	-0.127***	-0.0522	0.0932
	(0.0247)	(0.0342)	(0.0418)	(0.149)	(0.104)
Care	0.175***	0.193***	0.0896***	0.144	0.203***
	(0.0193)	(0.0287)	(0.0360)	(0.123)	(0.0827)
TV	0.108***	0.118***	0.0544	0.126	0.0792**
	(0.0149)	(0.0163)	(0.0578)	(0.142)	(0.0461)
N	32574	19943	4329	4205	1556

Standard errors in parentheses
Instrument = Grade within school average of race
* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.11: Send/Receive-Network Friendship Heterogeneity on GPA-OLS (NO CLUBS)

	All	White	Black	Hisp.	Asian
S/R. Hetero.	0.0141	-0.0626	0.211***	0.0763	-0.0847
	(0.0498)	(0.0698)	(0.0681)	(0.100)	(0.111)
Black	-0.165***				
	(0.0298)				
Hisp.	-0.183***				
	(0.0257)				
Asian	0.251***				
	(0.0362)				
Male	-0.166***	-0.171***	-0.194***	-0.115***	-0.184***
	(0.0193)	(0.0236)	(0.0276)	(0.0331)	(0.0415)
U.S.A.	-0.0836***	-0.0655	-0.137**	-0.0552*	-0.0891**
	(0.0250)	(0.0506)	(0.0769)	(0.0360)	(0.0517)
Coll. Mom	0.202***	0.236***	0.117***	0.102***	0.104***
	(0.0157)	(0.0180)	(0.0421)	(0.0309)	(0.0411)
Live Dad	0.130***	0.152***	0.145***	0.112***	0.208***
	(0.0148)	(0.0213)	(0.0309)	(0.0302)	(0.0604)
Health	0.170***	0.206***	0.0772***	0.120***	0.125***
	(0.0129)	(0.0158)	(0.0326)	(0.0339)	(0.0360)
Unsafe	-0.137***	-0.162***	-0.129***	-0.0756	0.0884
	(0.0232)	(0.0279)	(0.0346)	(0.0552)	(0.0940)
Care	0.176***	0.191***	0.0888***	0.126***	0.213***
	(0.0189)	(0.0255)	(0.0261)	(0.0577)	(0.0724)
TV	0.109***	0.119***	0.0356	0.0794**	0.111***
	(0.0146)	(0.0170)	(0.0526)	(0.0407)	(0.0398)
N	34509	20915	4626	4551	1702

Standard errors in parentheses

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.12: Send/Receive-Network Friendship Heterogeneity on GPA-IV (NO CLUBS)

	All	White	Black	Hisp.	Asian
S/R. Hetero.	0.204	1.715***	0.820	6.765	0.558
	(0.218)	(0.757)	(1.797)	(19.32)	(1.169)
Black	-0.179***				
	(0.0366)				
Hisp.	-0.225***				
	(0.0544)				
Asian	0.229***				
	(0.0437)				
Male	-0.166***	-0.164***	-0.206***	-0.0945	-0.193***
	(0.0192)	(0.0236)	(0.0421)	(0.0748)	(0.0415)
U.S.A.	-0.0869***	-0.0574	-0.118	-0.235	-0.129*
	(0.0248)	(0.0505)	(0.0902)	(0.520)	(0.0825)
Coll. Mom	0.202***	0.247***	0.110***	0.110***	0.0860*
	(0.0159)	(0.0191)	(0.0497)	(0.0467)	(0.0554)
Live Dad	0.130***	0.154***	0.144***	0.183	0.196***
	(0.0152)	(0.0273)	(0.0322)	(0.210)	(0.0656)
Health	0.170***	0.212***	0.0721**	0.126**	0.122***
	(0.0129)	(0.0169)	(0.0422)	(0.0643)	(0.0410)
Unsafe	-0.141***	-0.195***	-0.138***	-0.126	0.0874
	(0.0238)	(0.0291)	(0.0533)	(0.163)	(0.0919)
Care	0.177***	0.196***	0.0868***	0.107	0.202***
	(0.0191)	(0.0278)	(0.0295)	(0.106)	(0.0667)
TV	0.110***	0.127***	0.0356	0.164	0.107***
	(0.0150)	(0.0178)	(0.0543)	(0.277)	(0.0410)
N	34509	20915	4626	4551	1702

Standard errors in parentheses

Instrument = Grade within school average of race

* $p < .15$, ** $p < .10$, *** $p < .05$

Table A.13: First Stage Coefficients, IV Regressions-Race

	Instrument	All	White	Black	Hisp.	Asian
Same Race Fr.	Same Race Average	0.673*	0.573*	0.739*	1.056*	0.552
	(grade/school)	(0.0398)	(0.119)	(0.289)	(0.473)	(0.329)
R^2		0.296	0.0824	0.150	0.314	0.273
S. Hetero.	Same Race Average	-0.262*	-0.444*	-0.545*	-0.0797	-0.339
	(grade/school)	(0.0272)	(0.0419)	(0.182)	(0.117)	(0.187)
R^2		0.339	0.194	0.230	0.111	0.247
R. Hetero.	Same Race Average	-0.259*	-0.413*	-0.339*	0.0622	-0.528*
	(grade/school)	(0.0255)	(0.0594)	(0.175)	(0.166)	(0.137)
R^2		0.335	0.168	0.226	0.134	0.224
S/R. Hetero.	Same Race Average	-0.241*	-0.427*	-0.466 *	0.0540	-0.474*
	(grade/school)	(0.0280)	(0.0391)	(0.184)	(0.108)	(0.155)
R^2		0.368	0.230	0.280	0.190	0.247

Standard errors in parenthesis

* = Potential strong instrument

Table A.14: First Stage Coefficients, IV Regressions-Activities

	Instrument	All	Sport	Acad-Schol.
Same Act. Fr.	Same Activity Average	0.839*	0.857*	0.242
	(grade/school)	(0.0575)	(0.217)	(0.170)
R^2		0.0500	0.115	0.083
Common Club Fr.	Same Activity Average	0.344*	0.292	1.100*
	(grade/school)	(0.0508)	(0.245)	(0.158)
R^2		0.355	0.0714	0.0856

Standard errors in parenthesis

* = Potential strong instrument

Bibliography

Akerlof, G. and R. Kranton (2002), "Identity and Schooling: Some Lessons Learned for the Economics of Education ", *Journal of Economic Literature* 40: 1167-1201.

Alesina, A. and E. La Ferrara (2005), "Ethnic Diversity and Economic Performance", *Journal of Economic Literature* 43(3): 762-800.

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

Arcidiacono, P. , S. Khan, and J. Vigdor (2008), "Representation versus Assimilation: How do Preferences in College Admissions Affect Social Interactions?", Manuscript.

Arcidiacono, P. and A. Nathan (2007) "Descriptive and Structural Analysis of Racial and Academic Statistical Discrimination in Schools", Working Paper.

Arcidiacono, P. and J. Vigdor (2007) "Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College", Working Paper.

Bayer, P., H. Fang and R. McMillan (2005), "Separate When Equal? Racial Inequality and Residential Segregation", *National Bureau of Economic Research Working Paper # 11507.*

Bifulco, R., H. Ladd (2007), "School Choice, Racial Segregation, and Test Score Gaps: Evidence from North Carolina's Charter School Program", *Journal of Policy Analysis and Management* 26(1): 31-56.

Bowen, W. and D. Bok (1998) *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions* Princeton University Press.

Braddock III, J.H. (1981), "Race, Athletics, and Educational attainment: Dispelling the Myths", *Youth Society* 12: 335-350.

Brock, W. and S. Durlauf (2001), "Discrete Choice with Social Interactions", *Review of Economic Studies* 68: 235-259.

Broh, B. (2002), "Linking Extracurricular Programming to Academic Achievement: Who Benefits and Why?", *Sociology of Education* 75(1): 69-95.

Case, A. and A. Katz (1991), "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youths", *National Bureau of Economic Research Working Paper # 3705.*

Chantala, K. and J. Tabor (1999), "Strategies to Perform a Design-Based Analysis Using the Add Health Data", *Carolina Population Center-University of North Carolina at Chapel Hill*, Chapel Hill, NC.

Clotfelter, C. (2004) *After Brown: The Rise and Retreat of School Desegregation* Princeton University Press.

Drews, P. (2006) "The Effects of Desegregation and School Choice Programs on Student Achievement: An Examination of the Metco Program", Unpublished Manuscript.

Duncan, G., J. Boisjoly, M. Kremer, D. Levy and J. Eccles (2006), "Empathy or Antipathy? The Impact of Diversity", *The American Economic Review* 96(5): 1890-1905.

Foeman A. and T. Nance (1999), "From Miscegenation to Multiculturalism: Perceptions and Stages of Interracial Relationship Development", *Journal of Black Studies* 29: 540-557.

Foster, J.R. (2005), "Making Friends: A Non-Experimental Analysis of Social Group Formation", *Human Relations* 58: 1443-1465.

Gould, W., J. Pitblado and W. Sribney (2006) *Maximum Likelihood Estimation with Stata, third edition* Stata Press.

Hoxby, C. (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation", *National Bureau of Economic Research Working Paper # 7866.*

Kandel, D.B. (1978), "Homophily, Selection, and Socialization in Adolescent Friendships", *American Journal of Sociology* 84: 427-436.

Manski, C. (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies* 40: 531-542.

Marmaros, D. and B. Sacerdote (2006) "How Do Friendships Form?", *Quarterly Journal of Economics* 121(1): 79-119.

Mihaly, K. (2008) "Peer Effects of Actual Peers", Working Paper.

Moody, J. (2001), "Race, School Integration, and Friendship Segregation in America", *American Journal of Sociology* 107: 679-716.

Nehring, C. and C. Puppe (2002), "A Theory of Diversity", *Econometrica* 43(3): 762-800.

Raino, K. (1966) "A Study of Sociometric Group Structure: An Application of a Stochastic Theory of Social Interaction" in J. Berger, M. Zelditch, and B. Anderson, *Sociological Theories in Progress*, Vol.1, Houghton Mifflin.

Rumberger, R. and G. Palardy (2001), "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School", *Teachers College Record* 107: 1999-2045.

Tuma, N.B. and M. Hallinan (1979), "The Effects of Sex, Race and Achievement on Schoolchildren's Friendships", *Social Forces* 57: 1265-1285.

Udry, J. R. (2003), The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994-1996; Wave III, 2001-2002 [machine-readable data file and documentation], *Carolina Population Center, University of North Carolina at Chapel Hill*, Chapel Hill, NC.

Weinberg, B. (2006) "Social Interactions and Endogenous Association", Working Paper.

Xie, Y. and Z. Zeng (2002) "Statistical Models for Studying Intergroup Friendship", Annual Meeting of the American Sociological Association Methodology Section, Princeton, NJ.

Biography

I was born in Niagara Falls, NY on January 11th, 1978. I graduated from Duke University in 2000 with a B.S. degree in economics, a secondary major in mathematics, and a certificate in markets and management studies. I matriculated in 2002 to the economics PhD program at Duke University. I received an M.A. in economics in 2004. I have one co-authored publication, "Dementia and Medicare at Life's End" in the journal *Health Services Research*, co-authored with Dr. Frank Sloan and Dr. Vicki Lamb. I have received a pre-doctoral NIA fellowship in 2004 to study the economics of aging.