The Pennsylvania State University

The Graduate School

The College of the Liberal Arts

INTERSCHOLASTIC SPORTS PARTICIPATION AND ACADEMIC ACHIEVEMENT:

A NEW STATISTICAL APPROACH TO A CLASSIC QUESTION

A Thesis in

Sociology and Demography

by

Kristina L. Zeiser

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Arts

August 2008

The thesis of Kristina Zeiser was reviewed and approved* by the following:

George Farkas
Professor of Sociology and Demography
Thesis Adviser
Chair of the Graduate Program

Michelle L. Frisco
Assistant Professor of Sociology and Demography

Derek A. Kreager
Assistant Professor of Crime, Law, and Justice

*Signatures are on file in the Graduate School

Abstract

Research concerning the relationship between interscholastic sports participation and academic achievement has evolved over the past forty years. While early studies looked at cross-sectional data and community-based samples, more recent studies have used a longitudinal framework that controls for prior achievement. This study advances the field one step further by employing propensity score matching to find the effect of interscholastic and popular sports participation on student GPA in the twelfth grade. The findings of this study do not show evidence of the commonly-perceived positive impact that interscholastic sports have on educational outcomes. However, there is a positive effect on GPA for white (but not black) students who participate in popular sports (varsity basketball, football, baseball, and softball). Implications of the different operationalizations of sports participation and the different methods of analysis are discussed.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Sociological inquiry regarding the relationship between sports participation and academic outcomes dates back over forty years. Research concerning extracurricular activities began with Coleman's (1961) study of *The Adolescent Society*, and empirical studies followed with the intent to discover the effect of sports participation on academic and social outcomes. Since this seminal work, researchers have taken a great interest in students' athletic participation because of the unique status of the athlete in American society. Coleman's vision of the athlete was that of a member of the leading crowd, and the "athlete scholar" was ranked as the highest status in the school setting (Coleman 1961). More modern evaluations of the status of student athletes, which recognize the prominence of African Americans in sports, compare athletes to Greek gods, Dionysian in athletic skill and ability, but also violent and uncontrollable (Patterson 1998). While many adolescents aspire to be a star athlete, and many student athletes dream of careers as professional athletes, few are able to attain this goal. Accordingly, it is important that these student athletes achieve the education they will need to be successful in the labor market (Edwards 1984).

While it is commonly believed that sports activities are both academically and developmentally beneficial to youth, the empirical evidence is mixed. Thus, although several studies have shown that participation in athletic activities is associated with increased levels of academic achievement (Broh 2002; Eccles and Barber 1999; Feldman and Matjasko 2005) and educational attainment (Hanks and Eckland 1976; McNeal 1995), other studies have found that sports participation is not significantly related to grades in high school (Eitle and Eitle 2002; Hanks and Eckland 1976; Marsh 1993). Further, studies focusing on minority youth have found that sports participation does not significantly improve grades or test scores for Hispanic and

African American students (Eitle and Eitle 2002; Melnick, Sabo, and Vanfossen 1992). At least some of the variation in findings across studies appears to be due to the different data and methodologies that they have employed.

In this study, I focus on the effect of interscholastic sports participation on the academic outcome of grade point average (GPA) in the twelfth grade. While all academic and developmental outcomes are important for the individual's future success, GPA is a key outcome because higher GPAs have been shown to strongly increase an individual's likelihood of success in college and employment (Rosenbaum 2001). Test scores are important in that they assess students' cognitive ability, but these tests may have little bearing on the futures of students who intend to enter the workforce directly after high school. By contrast, GPA captures both the cognitive skills and the non-cognitive traits such as discipline and work ethic that are valued by educators and employers alike.

Prior studies have examined the relationship between sports participation and GPA, but not in the best possible way. In this study, I use a new statistical method-propensity score modeling-to improve on this prior research. I focus on each of two types of sports participation: participation in any interscholastic sport, and participation in a subset of very popular sports (varsity football, basketball, baseball, and softball). I also conduct separate analyses for white and black students, testing the claim that sports participation is particularly detrimental to African Americans (Eitle and Eitle 2002).

This study is organized as follows. First, I review prior theories and evidence on the relationship between sports participation and academic performance. Then, I describe propensity score modeling and explain the advantages of this methodology. Finally, I use both OLS regression and this new methodology analyze data from the first two waves of the Education

Longitudinal Study (ELS) of 2002 and 2004, estimating the effect of sports participation on GPA. The study concludes with a summary of the findings and a discussion of their implications.

PREVIOUS RESEARCH

Theoretical Perspectives

Past research has tested multiple theories that address the association between sports participation and academic outcomes. These theories view sports participation from one of two perspectives. The first views participation in sports as beneficial to students because it allows them the opportunity to develop their own identities and improve their aspirations for college. The second argues that sports participation may lower students' academic achievement because it directly detracts from the time needed to study for classes and instills an unrealistic aspiration to become a professional athlete. The following section provides an overview of the theories and perspectives that support these two points of view.

The positive effects of sports participation

A variety of theoretical perspectives have been employed to support a positive association between sports participation and academic success. Developmental theory (Broh 2002; Dworkin, Larson, and Hansen 2003; Schreiber and Chambers 2002) argues that students who participate in sports activities gain a unique opportunity to influence their own development based on their experience in the activity. Skills that are developed include goal-setting, time-management, deference to authority figures, control over impulsive behavior and emotions, and a deeper understanding of individuals from different backgrounds (Dworkin, Larson, and Hansen

2003). While this theory appears to support the idea that sports participation leads to positive academic outcomes, especially through the development of time-management and goal-setting behaviors, it has been suggested that athletes may choose to develop an identity that does not necessarily require academic success (Marsh 1992).

Social capital theory also suggests a positive relationship between sports participation and academic achievement. This theory emphasizes the use of sports activities to extend social networks and improve adolescents' ties with their parents, peers, teachers, school, and community (Broh 2002; Coleman 1988; Portes 1998). Interscholastic sports, which are normally organized by teachers or faculty within the school, provide a medium through which otherwise disengaged students may become integrated into the school community and obtain guidance toward succeeding in school. Coaches themselves are likely to be instrumental in helping athletes make decisions about college and their future careers (Snyder 1972). Adolescents' parents may also become more involved in the school community when their children participate in interscholastic sports, providing them with new networks of teachers, administrators, and other parents (Broh 2002). Through sports participation, disadvantaged students may obtain more knowledge about high school success and planning for college than they would otherwise have access to.

Two additional theories that emphasize the benefits of sports participation include the social control-integration theory (Hirschi 1969; McNeal 1995) and the leading-crowd hypothesis (Broh 2002; Coleman 1961). The social control-integration theory assumes that everyone is deviant by nature, but strong relationships with members of a community cause people to follow norms, fill prescribed roles, and avoid deviant behavior (Hirschi 1969). According to this theory, participation in sports activities prevents adolescents from partaking in deviant behaviors that are likely to influence their academic success (alcohol use, drug abuse, dropping out of school, etc.) due to strengthened ties with other students and school faculty. Similarly, the leading-crowd hypothesis assumes that students who are popular within the school setting are role models, and as a result they adopt positive academic behaviors and surround themselves with peers who also support positive academic behaviors (Coleman 1961). Both of these theories suggest that sports participation is positively related to academic success through the obligation that athletes feel to positively represent their team and school.

The negative effects of sports participation

Although some researchers believe that sports participation positively influences student academic outcomes, others have proposed that sports participation can actually be detrimental to students, and that this negative impact may be especially harmful to African American athletes (Patterson 1998). The zero-sum theory supports this negative view. This theory observes that time is finite, so that time spent in sports activities directly detracts from time spent on academic endeavors (Marsh 1992; Schreiber and Chambers 2002). In this view, athletes should have lower grades and test scores than non-athletes because time that should be reserved for studying and doing homework is instead taken up with the many hours of practice and performance that athletic activities entail.

While the zero-sum theory suggests that all athletes may suffer academically due to the time demands of sports activities, some researchers have suggested that sports may not be detrimental to all students, but rather they target minority students and exploit them for their athletic abilities (Edwards 1984). Further, it has been argued that participation in sports fosters unrealistic expectations for careers in the sports profession (Edwards 1984; Patterson 1998).

When students do not aspire to succeed in a field that requires an education, they are likely to minimize their concern for academic success at a young age. The lack of academic fervor in high school leads to under-preparation for college, and due to the low probability that college athletes will succeed in *either* their college courses *or* in the professional sports profession, these athletes are likely to experience the "treadmill to oblivion" as described by Harry Edwards:

> There is today disturbingly consistent evidence that the Black athlete who blindly sets out to fill the shoes of Dr. J., Reggie J., Magic J., Kareem Abdul-J., or O.J. is destined to end up with "No J."-no job whatsoever that he is qualified to do in our modern, technologically sophisticated society. (Edwards 1984, pg. 33)

Is there an effect of sports participation?

Each of these theories, whether predicting a positive or a negative association between sports and academics, posits that the relationship exists due to either internal influences (improved work-ethic, time management, educational expectations, maturation, etc.) or external relationships (with parents, coaches, the school, the community, their choice of friends, etc.). While it is probable that many of these factors partially explain the relationship between sports participation and academic outcomes, one final hypothesis must be considered: the selection hypothesis (Feldman and Matjasko 2005). Various studies have found that students who participate in high school sports systematically differ from non-participants: school athletes are more likely to be male and of a higher socioeconomic class, though participation rates do not appear to differ significantly by race (Fejgin 1994; Feldman and Matjasko 2005; Hanks and Eckland 1976; Marsh 1992; McNeal 1998). Also, students from smaller schools and private schools are more likely to participate due to the larger proportion of the student body needed to fill all of the positions on a team (Fejgin 1994; Holland and Andre 1987; McNeal 1998; McNeal 1999). According to the selection hypothesis, higher-achieving and more economically-advantaged students are more likely to participate in athletic activities, but sports participation does not affect academic outcomes above and beyond students' natural abilities. In other words, while research may find a positive association between sports participation and academic achievement, this may only be evident because the athletes were higher achievers before they joined the team.

Through the years, each of these theories have been tested empirically, but there is still little consensus regarding the nature of the relationship between sports participation and academic achievement. A review of the literature shows that the different findings were likely affected by the use of different datasets, methods, and operationalizations of both achievement and participation. The following section examines how the literature has developed and highlights conflicting findings across studies.

Empirical Results

Early empirical studies on the relationship between extracurricular activities and academic achievement were largely cross-sectional in nature, and many of these studies found that participation in interscholastic sports was associated with higher grades and higher grade point averages. For example, by matching athletes to non-athletes with similar background characteristics, Schafer and Armer (1968) found that athletes receive slightly higher grades than non-athletes. These authors suggest that improved relationships with teachers, coaches, and peers provide athletes with guidance and resources that foster academic success, a proposition that resembles social capital theory. During this time period, it was also reported that the athletes who were most likely to benefit academically from sports participation were those from disadvantaged, low-resource backgrounds (Phillips and Schafer 1971). Unlike recent reports that economically disadvantaged and minority athletes tend to place their athletic futures above their academic obligations, studies from this earlier time found that athletes have higher educational expectations than their non-athletic counterparts (Otto and Alwin 1977; Rehberg and Schafer 1968; Spreitzer and Pugh 1973), and this gives them more motivation to succeed academically. However, even during this time period, some studies found only a minimal or nonexistent relationship between sports participation and achievement (Hanks and Eckland 1976; Rehberg and Schafer 1968; Spreitzer and Pugh 1973).

As the passage of time brought with it new sampling and statistical techniques, these older studies have been labeled as inadequate or insufficient. First of all, due to the lack of available data, most of the studies from this time period looked at a snapshot of a community sample that was largely white and male (Otto and Alwin 1977; Spady 1970), and so many of these studies were not generalizable to the larger US population. Also, many of these studies were cross-sectional in nature, so researchers could not rule out the possibility that higher-achieving students may select themselves into sports (Phillips and Schafer 1971; Schafer and Armer 1968). The introduction of longitudinal data allowed researchers to more accurately estimate the effect of sports participation on future academic outcomes.

While nationally-representative, longitudinal datasets such as the High School and Beyond Survey (HS&B) or the National Education Longitudinal Study (NELS) of 1988 greatly benefited this research area, analyses of these new data continued to reach mixed findings. Using longitudinal data from the Michigan Study of Adolescent Life Transitions, Eccles and Barber (1999) found that students who participate in team sports have higher-than-expected GPAs in the twelfth grade. However, another study using more nationally-representative data found no significant relationship between sports participation and academic achievement (Schreiber and Chambers 2002).

Among studies analyzing longitudinal data, those that control for the measure of achievement at an earlier time point most accurately account for the possibility of positive selection into sports. This approach is the most common one taken by recent studies of sports participation and academic achievement. For example, analyzing the NELS dataset, Fejgin (1994) found that participation in sports activities was significantly associated with higher grades among 10^{th} graders after controlling these students' grades in the 8^{th} grade. Another study that utilizes the HS&B dataset found similar results: sports participation was positively related to senior grades after controlling for grades in the 10^{th} grade (Marsh 1992). These studies appear to rule out the selection hypothesis because they find that school athletes out-perform non-athletes academically *net of prior achievement*, and this prior achievement at least partially controls for many unobservable student traits such as cognitive skills, work ethic, discipline, and study habits that also affect achievement at the later point in time.

In a landmark study that utilizes the NELS dataset, Beckett Broh (2002) found that, of the numerous activities she included in her model, interscholastic sports have the strongest positive effect on educational achievement in the twelfth grade even after controlling for grades in the tenth grade. When a wide variety of activities (interscholastic sports, musical groups, vocational clubs, drama clubs, intramural sports, cheerleading, student council, and yearbook/journalism clubs) were included in the model simultaneously, participation in interscholastic sports was the only variable to remain significantly and positively related to both math and English grades. Moreover, further analysis showed that half of the effect of sports participation on grades was mediated through indicators of social capital (Broh 2002). Through the use of sound statistical methods, this study appears to provide not only definitive evidence of a positive relationship between sports participation and academic achievement, but also support for a theory that explains why this relationship exists.

While these studies utilized lagged regressor models and found positive effects of sports participation on academic outcomes, other studies have failed to find a significant relationship between participation and achievement for subgroups of the student population after controlling for prior achievement. For example, using the HS&B survey, Melnick et al. (1992) failed to find a significant effect of sports participation on grades and test scores for black and Hispanic students net of prior achievement. The only exceptions to these null findings were a negative effect on grades for suburban black boys and rural Hispanic girls and a positive effect on test scores for urban black boys and rural Hispanic girls (Melnick, Sabo, and Vanfossen 1992). Another study that utilized a lagged regressor model found that participation in interscholastic football and basketball is negatively associated with test scores for both black and white students, and while participating in these two sports does not significantly affect grades, participation in other interscholastic sports positively affects the grades of white athletes and negatively affects the grades of black athletes (Eitle and Eitle 2002). These results suggest that estimations of a *general* effect of sports participation on achievement may be inaccurate since the mechanisms through which participation affects academic achievement may operate differently by race.

In response to this continual inconsistency of findings, the current study introduces a new method of analysis by employing propensity score models to test the relationship between sports participation and GPA in the twelfth grade. Propensity score matching provides several improvements over OLS regression models controlling the lagged value of the dependent variable (GPA). First, it avoids the additive linear assumptions of OLS regression as well as OLS regression's tendency to extrapolate estimated effects outside the range (support) of data. While OLS regression models assume that the effect of sports participation and other predictors on achievement does not vary according to the values of other independent variables in the model, propensity score matching does not require this assumption since the difference in GPA outcomes between sports participants and non-participants is allowed to vary by respondents' propensity of participating in sports.

Thus, propensity score matching also allows the researcher to more accurately estimate the counterfactual, or the outcome that the athlete would have had if he or she had not participated on a sports team. Regression models are imperfect because the use of these models requires that one individual only belong to the "treatment" or the "control" group. The utilization of propensity score models make up for this shortcoming by estimating the counterfactual, which is calculated by comparing participants' outcomes with the outcomes of non-participants who are matched on a wide variety of observed characteristics. To demonstrate the advantages of propensity score modeling, the current study presents both regression and propensity score matching estimates, both using lagged variables. The following section describes propensity score modeling in more detail in order to reveal why this is a particularly useful method of analysis to test the relationship between sports participation and academic achievement.

PROPENSITY SCORE MODELING

This research utilizes propensity score modeling, a statistical technique that allows one to perform quasi-experimental analyses on survey data (Frisco, Muller, and Frank 2007;

Rosenbaum and Rubin 1983). Because this method attempts to approximate a randomized experiment using survey data, the language used to describe the components of propensity score models is the same language that is used in an experiment. For example, the independent variable of interest in a study is called the "treatment." The current study has two treatments: participation in interscholastic sports and participation in popular sports. Analysis of survey data requires that subjects are not randomly assigned to a treatment, but once individuals are matched on a substantial number of traits that are related to the treatment, the observation of who *actually receives* the treatment may be considered to be quasi-random. All of the respondents in a survey, then, either belong to the treatment group or the control group (did not participate in sports activities in the 10th grade).

The results of propensity score analysis are not necessarily the most accurate or best-conceived simply due to the matching of respondents. As with OLS regression, propensity score analyses cannot account for variables that are not included in the model, and so propensity score results are not immune to omitted variable bias (Rosenbaum and Rubin 1984). However, rather than simply controlling for background characteristics, in propensity score analysis, the effect of the treatment on the outcome is determined by comparing the outcomes, between those who did and did not receive the treatment, of those with similar *propensity scores*. Propensity scores are calculated using logistic regression or probit models that predict the probability of receiving the treatment. The independent variables within this probit or logistic regression equation must be associated with both the assignment of the treatment and the outcome. Based on the results of the logit (or probit) model, each respondent is assigned a propensity score which is defined as their probability of receiving the treatment. The selection of predictors that are included in the model ultimately determines the accuracy of the propensity score results. In general, the more predictors added to the model predicting the propensity of treatment, the more closely researchers may match treated and untreated cases.

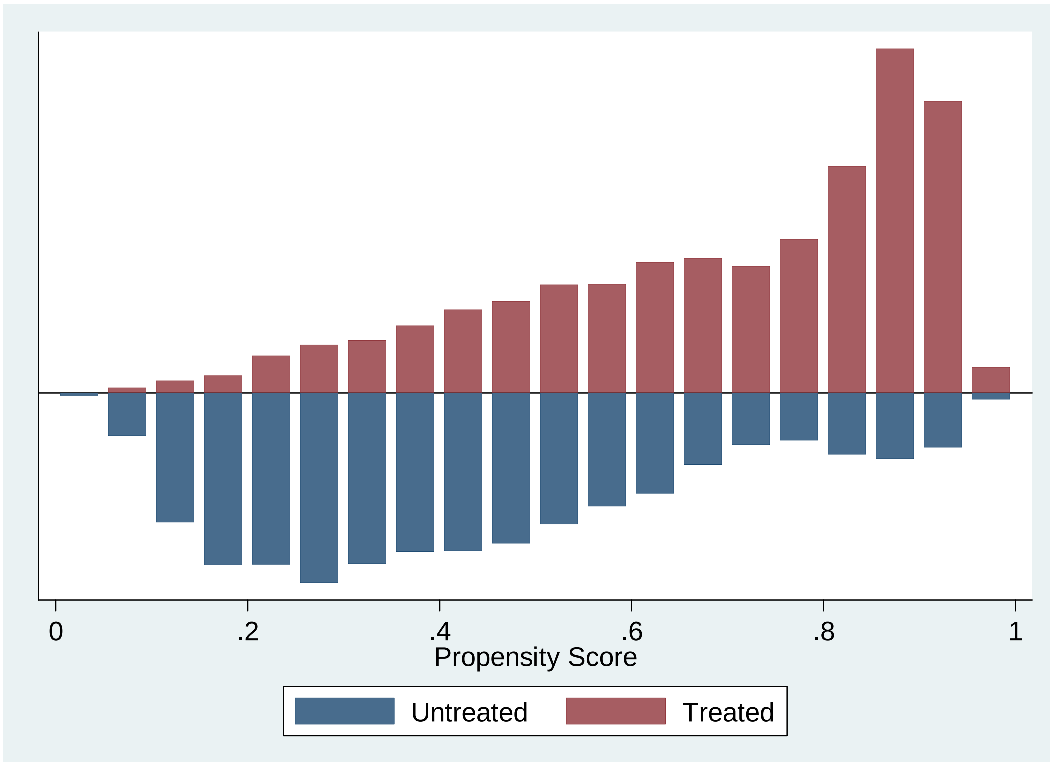
Once the probability of receiving the treatment is calculated, respondents are grouped into hierarchical strata (or blocks) based on their propensity scores. Quintiles are usually considered to provide sufficiently good matches, (Cochran 1968; Rosenbaum and Rubin 1984), and so most statistical packages will begin by grouping respondents into five strata. Next, t-tests are performed within each block to test whether the mean propensity scores of the treatment and control groups statistically differ. This is often referred to as “balancing” because this process makes sure that the propensity scores of treated cases are not significantly higher (or lower) than the propensity scores of control cases within a stratum. Assignment to the treatment cannot be considered quasi-random if only those with the highest propensity of receiving the treatment belong to the treatment group. In order to compare the outcomes of those who received and did not receive the treatment, the mean propensity scores of the treatment and control groups must be sufficiently similar within each stratum.

If the mean propensity score of the respondents who receive the treatment statistically differs from the mean propensity score of those who did not receive the treatment within a block, then the stratum is divided in half, and balancing resumes within each half of the stratum (Becker and Ichino 2002). When small sample sizes disallow further divisions of strata, then the model remains unbalanced. This would indicate that treatment and control groups cannot be compared because, based on the observed covariates in the model, respondents who received the treatment are not sufficiently similar to the respondents in the control group. When a stratum remains unbalanced, this issue may be addressed by adding more relevant predictors to the model predicting propensity scores. However, the balancing of propensity scores between the treatment and control groups is not guaranteed. For instance, balancing between treatment and control cases cannot occur if *all* of the treated cases have higher propensity scores than *all* of the members of the control group.

Once the mean propensity scores of the treated and control units are balanced within each stratum, then all of the covariates in the propensity score model should be checked to see if they balance between those who did and did not receive the treatment (Rosenbaum and Rubin 1983; Rosenbaum and Rubin 1984). In other words, among respondents with similar propensity scores (i.e., within each stratum), the mean values of the control variables should not significantly differ between the treated and control units. Interaction or higher-order terms may be added to the model in order to help balance the covariates (Rosenbaum and Rubin 1984), but then these terms must also be added to the model predicting the outcomes themselves. While it is not required that all variables balance within each stratum, having greater imbalance hints at the possibility of greater bias in the results (Hong and Raudenbush 2006).

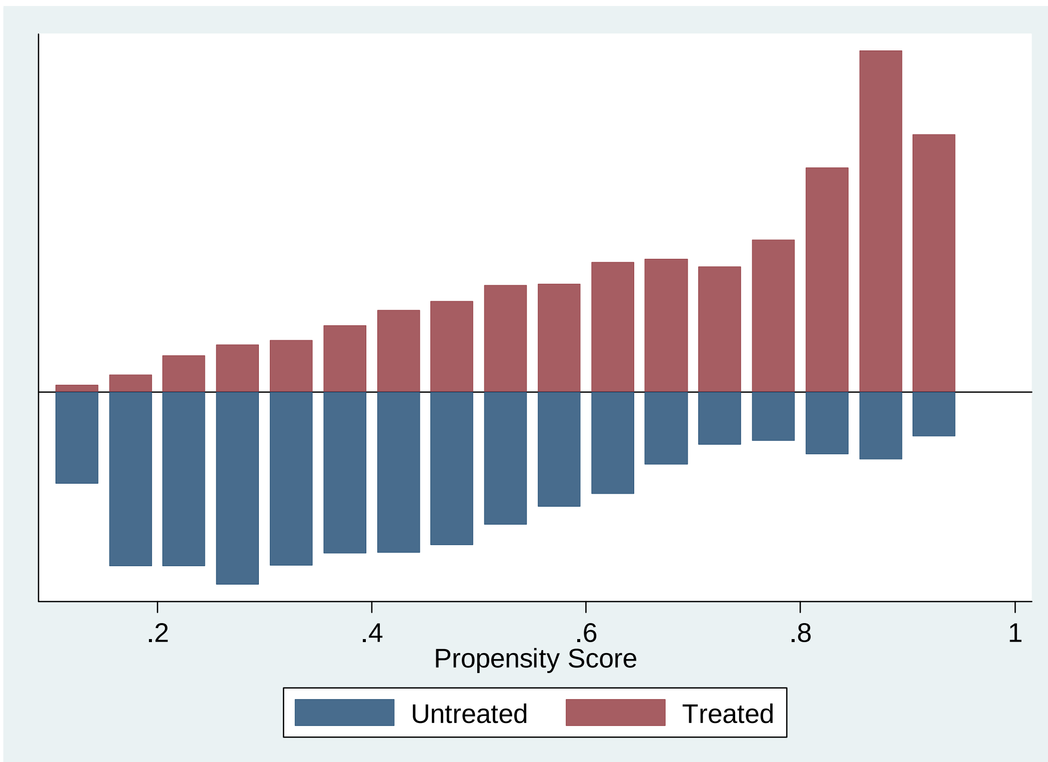
Once this stage of the process is complete, the results, termed the "average effect of treatment on the treated" or the ATT, may be calculated as the mean difference of the outcome between members of the treated and control groups with identical (or nearly-identical) propensity scores. This is not a simple calculation, however; researchers must ensure that their results are not unduly influenced by respondents with extreme values on the propensity score, and the ATT must be weighted to represent the population of respondents who receive the treatment. To clarify this issue, Figure 1 provides histograms of the propensity scores of treated and untreated respondents. This histogram represents the actual propensity scores calculated for interscholastic sports participation in this study. Figure 1 illustrates that respondents who participated in interscholastic sports are grouped at the higher end of the propensity score continuum while those who did not participate are largely found at the tail with the lower values.

Figure 1: Propensity Scores Distributions of Treated and Untreated Respondents



Estimations of the ATT often do not use every available case precisely because respondents who receive the treatment are not well matched to all respondents who do not receive the treatment. A "trimming" procedure is often performed to eliminate treatment or control cases that appear to be outliers based on their calculated propensity scores (Caliendo and Kopeinig 2005). For example, Figure 1 shows unmatched cases in the left-hand and right-hand tails of the distribution. The trimming procedure would eliminate some of these cases (usually one or two percent in each tail) to make sure that these outlying cases are not unduly biasing estimates of the ATT (Li and Zhao 2006; Smith and Todd 2005). Figure 2 illustrates what the distribution of propensity scores (by treatment) would look like if 2% of the treatment and control cases were removed from each of the tails.

Figure 2: Propensity Scores Distributions of Treated and Untreated Respondents after 2% Trimming



Another restriction that is often imposed when estimating the ATT is restricting analysis to the region of common support (Becker and Ichino 2002). To do this, control cases whose propensity scores do not reside between the minimum and maximum propensity scores of the treatment cases are removed from the sample. As with the trimming procedure, this eliminates outliers that may bias estimations of the ATT. Once any problematic outliers are removed from the sample, depending on the researcher's tolerance for the difference in propensity scores between treatment and control cases, the average effect of treatment on the treated may be calculated using four methods that use different matching techniques (Becker and Ichino 2002). If the method matches each treated case with more than one control case, weighting must be performed so that more weight is given to matches that have more similar propensity scores.

One of the most common matching techniques, called nearest neighbor matching, directly compares treated respondents with the control respondent that has the most similar propensity score (Becker and Ichino 2002; Caliendo and Kopeinig 2005). There is also an option to "allow for replacement." If the cases that received the treatment greatly outnumber the cases in the control group within a given range of propensity scores, it may be necessary to match a respondent who did not receive the treatment to more than one of the treated respondents (Caliendo and Kopeinig 2005). While this creates an oversample, by permitting some respondents to count more than once, allowing for replacement ensures that treated cases are not thrown out simply because there are not a sufficient number of control cases with the appropriate propensity score. Researchers have compared the results obtained through nearest neighbor matching both with and without allowing for replacement, and it was found that matching with replacement does not significantly bias the results (Dehejia and Wahba 1998). Once the difference in the outcome variable is measured within each matched pair, the ATT is calculated as the average difference in the outcome between the treated and control respondents. If the analysis allows for replacement, then the appropriate weights must be applied to the calculation of the ATT for those cases that are utilized more than once.

Another statistical strategy that is used to prevent the matching of treatment and control cases with greatly differing propensity scores is caliper matching. While matched cases should have identical or nearly-identical propensity scores, there is not a consensus over which value qualifies as an “acceptable” difference between propensity scores. For example, is a control unit with a propensity of 0.716 of receiving the treatment sufficiently similar to a treated case with a propensity score of 0.725? Caliper matching ensures that the nearest neighbor (as described by the previous method) is not a bad match. A bad match is defined as one in which, although the control respondent has the propensity score closest in value to the propensity score of the treated respondent, the difference between these two scores is too great to consider them to have similar propensities of receiving the treatment. Caliper matching imposes a restriction on the allowable difference in propensity scores between the treatment and control respondents (Caliendo and Kopeinig 2005), but because of the imposed limitations on what qualifies as a good match, fewer matches are likely to exist than those obtained through nearest neighbor matching.

One variation of this technique, called radius matching, takes into account the limits imposed by the caliper but then compares each treated respondent to all possible matches within the tolerable range of propensity scores. As with allowing for replacement, this creates an oversample because a treated case may be compared to many control cases, and so weights must be included in the estimation of the ATT to account for the cases that are used more than once. Caliper or radius matching may be preferred over nearest neighbor matching because the limitation of the caliper eliminates the chance of bad matches between the treatment and control groups. In caliper and radius matching, the average effect of treatment on the treated is calculated in the same way that it is calculated with nearest neighbor matching. The ATT is the average difference in outcome values between matched treatment and control respondents.

The next matching technique, stratification matching, retains respondents in hierarchical strata based on their propensity scores. Past research has shown that five strata are usually sufficient to remove most of the bias associated with the variables in the propensity score model (Cochrane and Chambers 1965; Imbens 2004). Through stratification matching, the mean difference of the outcome between the treatment and control respondents is calculated within each stratum rather than between a matched pair of respondents (Caliendo and Kopeinig 2005). The overall average effect of the treatment on the treated (ATT) for the sample is then found by weighting each block's mean difference (based on the number of cases within the stratum) and then finding the weighted average of the means (Becker and Ichino 2002).

Rather than selectively comparing members of the treatment and control groups, kernel matching weights the data and then matches every available pair in order to obtain the ATT (Becker and Ichino 2002). With kernel matching, the outcomes of each of the treated cases are compared to the outcomes of all of the members of the control group regardless of the propensity score. However, each difference is weighted based on the distance between the treated case's propensity score and the control case's propensity score (Becker and Ichino 2002; Caliendo and Kopeinig 2005). Weights are larger when the propensity scores are closer in value, and so the difference in the outcomes between members of a good match has greater weight in the calculation of the ATT than the difference in the outcomes between members of a bad match. This matching technique utilizes the strengths of the other techniques. Kernel matching both maximizes the number of cases used in the analysis and, to obtain the most accurate results, assigns weights to each pair of treated and control cases based on the similarity of their propensity scores.

Calculating estimates of the ATT using several of these techniques allows researchers to account for different obstacles presented by the data. Treatment and control cases may not be evenly distributed throughout the propensity score distribution (see Figure 1), and so the comparison of results obtained through the different matching methods provides stronger support for the results obtained. This study uses stratification, nearest neighbor, and kernel matching. Stratification matching will provide the average effect of treatment on the treated by grouping control and treatment cases with similar propensity scores. The results obtained by nearest neighbor matching will ensure that the outcome for each treated case is only compared to the outcome of the control case with the closest propensity score value. This will limit the sample size because control cases without matches will be eliminated from the sample, but it will also make sure that only cases with similar propensity scores (indicating that these respondents have similar background characteristics) are compared. This is especially useful in testing the selection hypothesis which suggests that athletes systematically differ from non-athletes on many background characteristics such as socioeconomic status, school type, and prior academic achievement (Fejgin 1994). The kernel matching technique will both increase the number of matches included in the analysis and ensure that the distance between the propensity scores of the treatment and control cases is taken into account. While all three of these techniques provide valid estimates of the effect of sports participation on twelfth grade GPA, any discrepancies between results may illustrate the incomparability of the control and treatment groups due to their differences in the observed covariates in the propensity score model.

<u>Data and Measures</u>

For the current study, I utilize the Education Longitudinal Study (ELS) datasets from 2002 (when respondents were in the 10^{th} grade) and 2004 (when respondents were in the 12^{th} grade). The data were collected for the National Center for Education Statistics (NCES) which is located within the United States Department of Education and the Institute of Education Sciences

(National Center for Education Statistics 2006). The ELS surveyed students, teachers, administrators, teachers, and librarians, and it also obtained the school records of the student respondents. The multi-stage sampling design first selected 750 schools throughout the United States and then randomly selected tenth graders within these schools, resulting in a sample size of over 15,000 students. The researchers oversampled non-public (private and Catholic) schools as well as racial minorities for comparison purposes (National Center for Education Statistics 2006).

This study's sample is limited to students who were not missing on any of the covariates in the propensity score model. This limits the sample to 9,760 respondents, or 59.6% of the Wave 1 sample. Much of the missingness (21% of the sample) is due to the attrition of respondents between Waves 1 and 2. While 20% of the sample is omitted due to missing values on the covariates, compared to a sample that is not missing on gender, race, socioeconomic status, or GPA in the tenth or twelfth grade (11,397), this study's sample does not significantly differ on these important demographic and academic characteristics. It is important to note that approximately 35% of this missingness is due to a large number of respondents (1,034) who do not have a measure of GPA for the 10^{th} grade, the 12^{th} grade, or both time points. One cause of this may be the fact that student GPA, this study's dependent variable, is taken from student transcripts.

The independent variables of interest in this study, or the treatments for which propensity scores are calculated, are participation in interscholastic sports and participation in popular sports. The variable for interscholastic sports participation is dichotomous with a value of 1 if the student participated in at least one of eight interscholastic sports activities in the tenth grade. These activities include football, soccer, baseball, softball, basketball, another individual sport, another team sport, or cheerleading/drill team. In this study's sample, 58.37% of the respondents participated in at least one interscholastic sport in the 10^{th} grade. Because such a large percentage of students participated in these sports activities, and previous studies have shown that the classification of activities matters when determining the effect of participation on academic outcomes (Broh 2002; Schreiber and Chambers 2002), separate analyses are run to determine the effect of participation in popular sports (varsity football, basketball, baseball, and softball) on GPA in the twelfth grade. In this sample, 17.35% of the respondents participated in these popular sports.

Individual-level variables in this study include both demographic and academic characteristics recorded during the tenth grade that are likely to influence both participation in sports activities and twelfth grade GPA. Included in the propensity score and OLS models are dummy variables indicating the race and family structure (single parent, cohabiting parent, or other household structure with married two-parent as the reference group) of the student, the socioeconomic status of the student's family, and whether English is a second language for the student. The variable for socioeconomic status is a standardized composite score created by the ELS that accounts for mother's education, father's education, family income, and both parents' occupations. This study also includes separate measures of mother's and father's level of education (in years of schooling) and whether the mother or the father is unemployed at the time of the 10^{th} grade survey.

Other activities that occupy students' leisure time are measured with variables indicating participation in intramural sports, membership in non-athletic clubs, and the extent to which students participate in non-school-related sports. The measures for intramural sports participation and participation in non-athletic clubs are dichotomous with a value of 1 indicating that the student participates in at least one intramural sport or club. The types of non-athletic clubs considered in this study include band or music-related, drama, national honors society, yearbook or school newspaper, service, academic, hobby, and vocational. The indicator for students' participation in non-school sports ranges from 0 (never plays non-school sports) to 3 (plays a non-school sport every day or almost every day). This variable serves as a proxy measure for both students' non-school-related use of leisure time and students' athletic inclinations.

Variables related to students' educational backgrounds are also included in the propensity score and OLS models. These variables include students' academic track (college prep or vocational with academic track as the reference group), educational expectations, and reading and math test scores from tests administered by the data collectors. Student educational expectations is measured with five dummy variables with "obtaining a high school diploma or a GED" as the reference category: not completing high school or a GED, receiving some college education, receiving a four-year degree, receiving more than a four-year degree, and "don't know." Because the dependent variable in this study is GPA in the 12th grade, this study attempts to account for any unobserved student characteristics by controlling for GPA in the 10th grade (also taken from the students' high school transcripts).

Due to the potential influence of school characteristics on student athletic participation and achievement, dummy variables are included in the analyses that indicate whether the school is located in an urban or rural area (suburban reference) and whether the student attends a private or Catholic school (with public school as the reference group). Due to a large amount of missing data on other school-level variables, no additional school-level characteristics are included in the analyses, but the effect that these variables would have on 12th grade GPA should be accounted for by controlling for 10th grade GPA.

Methods of Analysis

This study performs both propensity score analysis and OLS regression to determine the effect of sports participation on 12th grade GPA. Two types of sports participation are investigated in the current study: participation in interscholastic sports (any sport) and participation in popular sports (varsity football, basketball, baseball, or softball). Analyses are performed three times for each treatment: once for the entire sample, once for white students only, and once for black students only. As a result, propensity sores, the ATT, and the OLS results are each estimated six separate times.

The use of propensity score models can be described as a two-stage process. First, propensity scores are estimated using logistic regressions that predict respondents' probability of participating in interscholastic or popular sports. During this stage, it is important to note how well the propensity scores and covariates balance within strata between those who participated in the sports activity and those who did not. After the estimation of the propensity scores, respondents with propensity score values below the second percentile and above the 98th percentile are "trimmed" or removed from the sample because they may potentially bias the estimates of the ATT (for another study that uses a similar trimming strategy, see Li and Zhao 2006). Analyses are also restricted to respondents within the region of common support, and so non-participants whose propensity scores do not fall between the minimum and maximum propensity scores of participants are also removed from the sample. To conserve space, tables that report the descriptive statistics about the strata (after trimming) will be provided in

Appendix B. In these tables, the number of cases, as well as the means and standard deviations of propensity scores within each stratum, will be reported separately for sports participants and non-participants.

The second stage of propensity score analysis involves the estimation of the average effect of treatment on the treated. Participants and non-participants are matched based on their estimated propensity scores, and the average effect of sports participation on 12^{th} grade GPA is calculated using three different matching strategies: stratification, nearest neighbor, and kernel matching. Supplemental analyses were performed that imposed a more stringent trimming procedure, removing respondents with propensity scores below the fifth percentile and above the ninety-fifth percentile, and while the general trend of results does not change, it is noted when differences in results arise. The results of the propensity score analysis are novel in that they use a quasi-experimental research design to model the relationship between sports participation and student GPA.

Finally, the estimates of the ATT are compared with estimates obtained through OLS regressions predicting 12^{th} grade GPA while controlling for GPA in the 10^{th} grade. While it may be argued that predicting a change-score may more accurately model the effect of sports participation on changes in achievement over time, Paul Allison (1990) argues that, for situations in which the "treatment" may be associated with the time 1 measure of the dependent variable, lagged dependent variable analysis may be preferred. Because many studies have indeed found a positive association between sports participation and student GPA, I use the lagged dependent variable specification. These regression models control for every covariate and interaction term that is present in the propensity score model, and both models control for the lagged dependent variable (10th grade GPA), so that the results from the OLS models are directly comparable to the results obtained through the propensity score analysis.

RESULTS

Interscholastic Sports Participation

Descriptive Statistics

To begin, the means and proportions of all of the variables used in the OLS regressions and the propensity score models are shown in Appendix A Table A1. These descriptive statistics are provided for the entire sample and separately for interscholastic sports participants and non-participants. Table A1 also shows the differences between participants and non-participants on all of these characteristics and the significance of these differences. The descriptive statistics show which student characteristics may be particularly important when determining one's propensity of participating in interscholastic sports. According to these results, students who participate in interscholastic sports generally experience more educational success as evidenced by higher 10th and 12th grade GPAs, higher reading and math test scores, and a larger proportion of students in the college prep academic track. Interscholastic athletes also appear to have more privileged economic backgrounds with a higher average of years of parental education and higher parental socioeconomic status than non-participants. Sports participants are also more likely than non-participants to participate in other extra-curricular activities and have a smaller percentage of students that are female, Asian, or Hispanic. Other characteristics that appear to be important determinants of interscholastic sports participation include family structure (with smaller proportions of participants belonging to non-traditional family structures) and school type. These results indicate that students who participate in interscholastic sports differ from non-participants on many educational, demographic, and school characteristics.

Estimates for the Entire Sample

The first goal of this paper is to determine whether there is an overall effect of interscholastic sports participation on 12th grade GPA. To calculate the propensity score, all of the background characteristics in Table A1 are included in a logistic regression model predicting interscholastic sports participation. When utilizing the entire sample that is not missing values on any of the covariates, eleven strata are created in which the propensity scores and background characteristics balance between participants and non-participants (see Appendix B, Table B1 for stratum frequencies, means, and standard deviations). After observing which covariates do not balance between participants and non-participants within each stratum, an interaction term between intramural participation and reading test scores is included in the propensity score model in order to minimize the number of covariates that do not balance. After adding this interaction term, only four covariates do not balance between interscholastic sports participants and non-participants: the race dummy variable for black students does not balance in the second stratum, the dummy variable for cohabiting parent structure does not balance in the third stratum, the dummy variable for “other parent structure” does not balance in the sixth stratum, and the variable for reading test scores does not balance in the seventh stratum. After performing 2% trimming and imposing the constraint of common support, the final sample includes 9,360 respondents, 5,512 of whom are interscholastic sports participants.

Estimates of the average effect of treatment on the treated are calculated using the stratification, nearest neighbor, and kernel matching techniques. All of these results, which are presented in Table 1, indicate that there is not a significant effect of interscholastic sports participation on 12th grade GPA once participants and non-participants are matched on the observed characteristics included in this study. The nearest neighbor matching technique uses the smallest number of “control cases” (non-participants) because it only compares participants to the single non-participant with the closest propensity score, and each control case may be matched with more than one treatment case. This matching technique calculates an ATT that is essentially 0. While the stratification and kernel matching techniques calculate a small, positive ATT, indicating that sports participants have higher 12th grade GPAs when matched with non-participants with similar propensity scores, none of these estimates achieve statistical significance.

Table 1: Estimates of the Effect of Interscholastic Sports Participation on 12th Grade GPA

	Interscholastic Participation			N Treatment	N Control	Total N
Propensity Score Matching						
Stratification	0.018		(0.021)	5512	3848	9360
Nearest Neighbor	0.000		(0.028)	5512	1804	7316
Kernel Matching	0.019		(0.020)	5512	3848	9360
OLS Regression						
Estimate	0.030	**	(0.012)			9360

Unlike the propensity score results, the results from the OLS regression show that there is a small positive effect of interscholastic sports participation on 12th grade GPA that is significant at the 0.01 level. It is important to note that this OLS regression includes all of the variables in the propensity score model (including the interaction between intramural participation and reading test scores) so that the results may be directly compared. Based on the OLS estimate, on average, sports participants have 12th grade GPAs that are 0.030 points higher than the GPAs of non-participants even after controlling for 10th grade GPA. By comparison, the standard deviation of 12th grade GPA for this sample is 0.73. It is particularly interesting in this study that, despite the fact that they use the same sample and covariates, the OLS estimate and the estimates of the ATT present conflicting results.

Estimates by Race

Past research has suggested that sports participation may affect students differently depending on their race (Sabo, Vanfossen, and Melnick 1993). To determine whether the effect of interscholastic sports participation on 12th grade GPA varies by students' race, propensity scores are estimated separately for black and white students. To estimate the propensity scores, all of the independent variables in Table A1 (excluding the dummy variables for race) are included in two logistic regression models predicting interscholastic sports participation. Neither of the logistic regression equations requires interaction or higher-order terms in order to balance the propensity scores within the strata.

For white respondents, the calculated propensity scores balance within nine strata (see Appendix B, Table B2 for stratum frequencies, means, and standard deviations). Within these strata, all of the covariates balance between interscholastic sports participants and non-participants with the exception of math test scores in the first stratum, reading test scores in the fourth stratum, and the dummy variable for attending an urban school in the fifth stratum. After performing a 2% trimming procedure and eliminating cases that do not belong to the region of common support, the ATT is estimated using a sample of 5,750 white respondents, 3,649 of whom participated in interscholastic sports in the tenth grade.

Propensity scores balance within six strata for African American respondents (see Appendix B, Table B3 for stratum frequencies, means, and standard deviations). Within each stratum, each of the covariates balance between the treatment and control groups except for reading and math test scores in the second stratum. After performing the 2% trimming procedure and imposing the constraint of common support to remove outliers, the final sample includes 1,013 black students, and 575 of them participated in interscholastic sports in the 10th grade.

Estimations of the effect if interscholastic sports participation on 12th grade GPA for white and black students are provided in Table 2. The propensity score results indicate that sports participation does not have a significant effect on 12th grade achievement for either white or black students. For white students, all of the matching strategies estimate a small positive effect of interscholastic sports participation on 12th grade GPA, but none of these estimates achieve statistical significance. For black students, the results estimated using stratification and kernel matching techniques indicate that sports participants may have *lower* twelfth grade GPAs than black non-participants, but neither of these estimates achieve statistical significance.

Table 2: Estimates of the Effect of Interscholastic Sports Participation on 12th Grade GPA by Race

	Interscholastic Participation			N Treatment	N Control	Total N
White Particiants						
Propensity Score Matching						
Stratification	0.003		(0.024)	3649	2101	5750
Nearest Neighbor	0.014		(0.032)	3649	1098	4747
Kernel Matching	0.005		(0.024)	3649	2101	5750
OLS Regression						
Estimate	0.029	*	(0.015)			5750
Black Participants						
Propensity Score Matching						
Stratification	-0.021		(0.070)	575	438	1013
Nearest Neighbor	0.031		(0.089)	575	200	775
Kernel Matching	-0.031		(0.071)	575	438	1013
OLS Regression						
Estimate	0.019		(0.044)			1013

The results of the OLS regressions, on the other hand, indicate that sports participation has a significant effect on twelfth grade GPA for white students but no effect on the GPAs of black students. As with the estimates that utilize the entire sample, the OLS regression estimate for whites indicates a small positive effect of sports participation on 12th grade GPA even after controlling for 10th grade achievement, but the propensity score estimates fail to find a significant relationship. It should be noted, however, that after performing the more stringent 5% trimming procedure, this estimate is no longer statistically significant.

Participation in Popular Sports

Descriptive Statistics

The second analytical goal of this study is to determine whether participation in popular sports significantly affects student GPA in the 12th grade and whether this relationship varies by race. Students who participate in their high school's varsity baseball, softball, football, or basketball teams in the 10th grade are classified as popular sports participants or members of the "treatment" group. The means and proportions of the independent variables that are included in the propensity score and OLS models are provided by treatment status in Appendix A Table A2.

From these descriptive statistics, it is apparent that interscholastic sports participants and popular sports participants do not have the same characteristics. While interscholastic athletes in general evidence more academic success than non-athletes, popular sports participants actually have *lower* reading test scores on average compared to non-participants, and their math test scores and 10th grade GPAs do not significantly differ from non-participants. However, popular sports participants do have higher average GPAs in the 12th grade and are less likely to be in the vocational academic track than non-participants. Participants in popular sports have higher average parental SES than non-participants, but, unlike interscholastic athletes, father's education and family structure do not significantly differ by participation status. Athletes in popular sports are less likely to be female, Asian, and Hispanic than non-participants, but black students are overrepresented within these sports. Finally, students who attend private and rural schools are more likely to participate in popular sports while students in urban schools are less likely to participate in these sports. Overall, it is clear that many demographic, academic, and school-level variables significantly differ between participants and non-participants of popular sports.

Estimates for the Entire Sample

To calculate propensity scores for the entire sample, all of the respondents' family, education, and school characteristics in Table A2 are included in a logistic regression predicting individuals' popular sports participation. After investigating which covariates do not balance between the treatment and control groups within each stratum, interaction terms are added to the propensity score equation that interact family structure (cohabiting parents, single parent, or other family structure) with school type (catholic or private school) to minimize the number of covariates that do not achieve balance.

Once these interaction terms are added to the propensity score model, the propensity scores balance within nine strata, and the only covariates that do not balance between popular sports participants and non-participants are mothers' education in the second stratum, math test scores in the fourth stratum, and 10^{th} grade GPA in the eighth stratum. After the 2% trimming procedure is performed and the sample is limited to respondents who lie within the region of common support, estimates of the ATT utilize a sample of 9,365 respondents, 1,596 of whom participated in popular sports in the tenth grade (see Appendix B, Table B4 for stratum frequencies, means, and standard deviations).

Based on the estimations of the average effect of treatment on the treated, participation in popular sports positively affects students' 12th grade GPAs even after matching participants to non-participants with very similar background characteristics. The ATT obtained through nearest neighbor matching indicates that popular sports athletes have 12th grade GPAs that are an average of 0.072 points higher than non-participants, and this result is significant at the 0.05 confidence level (Table 3). While this is the largest effect estimated through propensity score matching, this effect is relatively small because it is approximately one-tenth of a standard deviation in 12th grade GPA. The results obtained through stratification and kernel matching also estimate small positive effects of popular sports participation on 12th grade GPA, and these estimates achieve significance at the 0.01 confidence level. This is likely to be a result of the larger number of control cases that these techniques utilize in order to calculate the ATT. Interestingly, the results from the OLS regression indicate an average effect of popular sports participation that is very similar to the estimate predicted by stratification and kernel matching.

Table 3: Predicting the Effect of Popular Sports Participation on 12th Grade GPA

	Popular Sport Participation			N Treatment	N Control	Total N
Propensity Score Matching						
Stratification	0.057	**	(0.021)	1596	7769	9365
Nearest Neighbor	0.072	*	(0.029)	1596	1286	2882
Kernel Matching	0.049	**	(0.021)	1596	7769	9365
OLS Regression						
Estimate	0.053	***	(0.015)			9365

Estimates by Race

To determine whether the effect of popular sports participation on 12th grade GPA differs by race, propensity scores are estimated separately for black and white students. To calculate propensity scores, all of the covariates in Table A2 (excluding the dummy variables for race) are included in two logistic regression models predicting participation in popular sports. The balance of propensity scores and most of the covariates between treatment and control cases within each stratum is achieved without the addition of interaction or higher-order terms for both white and black students.

For white students, all of the covariates in the propensity score model balance between participants and non-participants within eight strata with the exception of parents' SES in the first stratum and a dummy variable for student educational expectation (completing a four-year degree) in the eighth stratum (see Appendix B, Table B5 for stratum frequencies, means, and standard deviations). After performing the trimming procedures, the final sample includes 5,746 white respondents, and 1,067 of these respondents participated in popular sports in the tenth grade.

For African American students, the calculated propensity scores balance within six strata (see Appendix B, Table B6 for stratum frequencies, means, and standard deviations). All of the covariates also balance within these strata with the exception of math test scores in the sixth stratum and a dummy variable indicating that the student is unsure about his or her academic future (responded "don't know" to the question pertaining to educational expectations) in the second stratum. After performing the trimming procedures, the remaining sample includes 973 African American respondents, 245 of whom participated in popular sports in the tenth grade.

The results of the propensity score analysis indicate that popular sports participation positively influences 12th grade GPA, but only for white students. The estimated effect of popular sports participation for white students ranges between 0.067 and 0.074 depending on the matching strategy, a value that approximates one-tenth of a standard deviation in 12th grade GPA (Table 4). While all of these estimates are significant at the 0.01 confidence level, the ATT obtained through nearest neighbor matching loses significance when the 5% trimming procedure is performed. The results from the OLS regression also indicate a positive relationship between popular sports participation and 12th grade GPA for whites, and the size of the estimated coefficient is very similar to the estimates of the ATT.

Table 4: Estimates of the Effect of Popular Sports Participation on 12th Grade GPA by Race

	Popular Sport Participation			N Treatment	N Control	Total N
White Particiants						
Propensity Score Matching						
Stratification	0.068	**	(0.023)	1067	4679	5746
Nearest Neighbor	0.074	*	(0.032)	1067	880	1947
Kernel Matching	0.067	**	(0.021)	1067	4679	5746
OLS Regression						
Estimate	0.061	***	(0.017)			5746
Black Participants						
Propensity Score Matching						
Stratification	0.004		(0.059)	245	728	973
Nearest Neighbor	0.060		(0.079)	245	181	426
Kernel Matching	0.019		(0.063)	245	728	973
OLS Regression						
Estimate	0.046		(0.047)			973

While a positive association is evident for white students, these results indicate that there is not a significant relationship between participation in popular sports and 12th grade GPA for black students. While it is possible that the positive estimates do not achieve statistical significance due to the smaller number of black respondents, the effect sizes themselves are somewhat smaller for black students compared to white students. Also, despite theories that would predict that sports participation leads to lower levels of academic achievement for African American students, these results do not find evidence of a negative relationship. Instead, these results indicate that the widely-perceived positive effect of sports participation on school achievement may only exist for white athletes that participate in popular sports.

DISCUSSION

The purpose of this study was twofold. First, through the use of propensity score modeling, this study attempted to clarify the relationship between sports participation and student achievement. Second, I investigated whether this relationship differs by race. The results of the propensity score models indicate that participation in popular sports, but not interscholastic sports in general, has a significant positive effect on 12^{th} grade GPA even after matching respondents based on various background characteristics. While this positive effect was evident for white students, the effect of popular sports on 12^{th} grade GPA failed to achieve significance for African American students. It must be recognized, however, that this study's sample contains a relatively small proportion of black students, and so it is possible that the different effects by race may be an artifact of differing sample sizes.

Another potential source of error in this study lies in the estimation of the propensity scores themselves. The propensity score estimates rely on the assumption that all relevant variables that predict the treatment are included in the model. Survey research is always burdened with the problem of omitted or unobserved variable bias, but this study attempts to minimize this limitation by including an earlier measure of the dependent variable in the model.

By controlling for students' 10th grade GPA, calculations of the ATT rely on the comparison of students who are already matched based on their previous levels of educational achievement, and so this should minimize the extent to which unobserved characteristics are biasing the estimates.

Despite these potential limitations, the conflicting findings within this study highlight the areas of research that require further analysis. For example, the fact that popular sports participation significantly affects 12th grade GPA while participation in interscholastic sports in general does not suggests that there may be qualitative differences between these two types of athletic activities. It is evident from the descriptive statistics of interscholastic athletes and non-participants that these athletes generally come from higher-SES backgrounds and have higher levels of academic achievement in the tenth grade. This provides preliminary support for the selection hypothesis because it appears that students who have more advantaged backgrounds and already have higher levels of achievement choose to participate in sports, and it is not participation in the sport itself that improves athletes' level of achievement.

Popular sports, however, do not experience the same positive selection into participation. Try-outs are very rigorous for sports such as football and basketball, especially for tenth graders participating on the varsity team, and so factors such as test scores and parents' education may not be important determinants of participation on these teams. This is evident in Table A2 where popular sports participants do not differ from non-participants on many indicators of educational background and family life. It is likely that higher-achieving and economically advantaged students join interscholastic teams because they have positive associations with school in general, or in order to appear well-rounded on college applications, but participation on popular varsity teams is largely restrictive and impervious to such "supply-side" selection effects. Future research about the effects of sports participation on student outcomes should recognize the possibility that students and their parents may play a large role in the decision to participate in some sports, but they may not have much power over the participation on certain teams that may be more selective on athletic ability alone.

The contradictory results regarding the relationship between interscholastic sports participation and 12th grade GPA are also important for the field. The OLS regression estimates for the entire sample and for white students indicate that interscholastic sports participation leads to higher GPAs in the twelfth grade, but estimates of the ATT using propensity score modeling fail to find a significant relationship. I argue that the propensity score results more accurately model this relationship because, while OLS regression produces an estimate controlling for background characteristics, propensity score models do not make the assumption that the distribution of these control variables is the same for participants and non-participants. Propensity score models match respondents based on their propensity scores to ensure that estimates are not calculated through the comparison of participants and non-participants that are qualitatively different from one another. The trimming procedure and the region of common support are two examples of the unique features that propensity score modeling has to offer to ensure that treatment effects are not biased by extreme cases on either end of the propensity score continuum.

Finally, this study supports the literature pertaining to the different effects of sports participation on black and white students. Some researchers have recognized sports to be an outlet through which under-privileged minorities may gain access to educational resources and social ties that may help them to succeed academically. Other researchers have suggested that sports participation hurts African American students by instilling in them false hopes of athletic careers and robbing them of the time and energy needed to succeed academically. The results of this study do not support either of these theories. However, while sports participation was not related to *lower* academic achievement for African American students, it must be recognized that this group of students appears to be denied the benefits that participation in popular sports provides to white students. Further research should seek to determine the mechanisms through which participation in popular sports positively affects the achievement of some students while providing zero benefits to other students. If Americans are to continue idolizing the image of the athlete and encouraging their peers and children to participate, then they need to make sure that these athletes are adequately prepared for whatever their futures hold.

Work Cited

Allison, Paul D. 1990. "Change Scores as Dependent Variables in Regression Analysis." *Sociological Methodology* 20:93-114.

Becker, Sascha O. and Andrea Ichino. 2002. "Estimation of Average Treatment Effects Based on Propensity Scores." *The Stata Journal* 2:358-377.

Broh, Beckett A. 2002. "Linking Extracurricular Programming to Academic Achievement: Who Benefits and Why?" *Sociology of Education* 75:69-91.

Caliendo, Marco and Sabine Kopeinig. 2005. *Some Practical Guidance for the Implementation of Propensity Score Matching*: IZA: Institute for the Study of Labor.

Cochran, William G. 1968. "The Effectiveess of Adjustment by Subclassification in Removing Bias in Observational Studies." *Biometrics* 24:205-213.

Cochrane, W G. and S. Paul Chambers. 1965. "The Planning of Observational Studies of Human Populations." *Journal of the Royal Statistical Society, Series A (General)* 128:234-266.

Coleman, James S. 1961. *The Adolescent Society*. Glencoe IL: Free Press.

—. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S94-S120.

Dehejia, Rajeev H. and Sadek Wahba. 1998. *Propensity Score Matching Methods for Non-experimental causal studies*. Cambridge: National Bureau of Economic Research.

Dworkin, JB, R Larson, and D Hansen. 2003. "Adolescents' Accounts of Growth Experiences in Youth Activities." *Journal of Youth and Adolescence* 32:17-26.

Eccles, JS and BL Barber. 1999. "Student Council, Volunteering, Basketball, or Marching Band: What Kind of Extracurricular Involvement Matters?" *Journal of Adolescent Research* 14:10-43.

Edwards, Harry. 1984. "The Collegiate Athletics Arms Race: Origins and Implications of the "Rule 48" Controversy." in *Fractured Focus*.

Eitle, Tamela McNulty and David J. Eitle. 2002. "Race, Cultural Capital, and the Educational Effects of Participation in Sports." *Sociology of Education* 75:123-146.

Fejgin, Naomi. 1994. "Participation in High School Competitive Sports: A Subversion of School Mission or Contribution to Academic Goals?" *Sociology of Sport Journal* 11:211-230.

Feldman, A F and J L Matjasko. 2005. "The Role of School-Based Extracurricular Activities in Adolescent Development: A Comprehensive Review and Future Directions." *Review of Educational Research* 75:159-210.

Frisco, Michelle L., Chandra Muller, and Kenneth Frank. 2007. "Parents' Union Dissolution and Adolescents' School Performance: Comparing Methodological Approaches." *Journal of Marriage and Family* 69:721-741.

Hanks, M and B Eckland. 1976. "Athletics and Social Participation in the Educational Attainment Process." *Sociology of Education* 49.

Hirschi, Travis. 1969. *Causes of Delinquency*. Los Angeles: University of California Press.

Holland, Alyce and Thomas Andre. 1987. "Participation inExtracurricular Activities in Secondary School: What is Known, What Needs to be Known." *Review of Educational Research* 57:437-466.

Hong, Guanglei and Stephen W. Raudenbush. 2006. "Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data." *Journal of the American Statistical Association* 101:901-910.

Imbens, Guido. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *The Review of Economics and Statistics* 86:4-29.
Li, Xianghong and Xinlei Zhao. 2006. "Propensity Score Matching and Abnormal Performance after Seasoned Equity Offerings." *Journal of Empirical Finance* 13:351-370.
Marsh, Herbert W. 1992. "Extracurricular Activities: Beneficial Extension of the Traditional Curriculum or Subversion of Academic Goals?" *Journal of Educational Psychology* 84:553-562.
—. 1993. "The Effects of Participation in Sport During the Last Two Years of High School." *Sociology of Sport Journal* 10:18-43.
McNeal, Ralph B Jr. 1995. "Extracurricular Activities and High School Dropouts." *Sociology of Education* 68:62-81.
—. 1998. "High School Extracurricular Activities: Closed Structures and Stratifying Patterns of Participation." *Journal of Education Research* 91:183-191.
—. 1999. "Participation in High School Extracurricular Activities: Investigating School Effects." *Social Science Quarterly* 80:291-309.
Melnick, MJ, DF Sabo, and BE Vanfossen. 1992. "Educational Effects of Interscholastic Athletic Participation on African American and HIspanic Youth." *Adolescence* 27:295-308.
Otto, LB and DF Alwin. 1977. "Athletics, Aspirations, and Attainments." *Sociology of Education* 42.
Patterson, Orlando. 1998. *Rituals of Blood: Consequences of Slavery in Two American Centuries*. New York: Basic Civitas.
Phillips, John C. and Walter E. Schafer. 1971. "Consequences of Participation in Interscholastic Sports: A Review and Prospectus." *The Pacific Sociological Review* 14:328-338.
Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24:1-24.
Rehberg, RA and WE Schafer. 1968. "Participation in Interscholastic Athletics and College Expectations." *American Journal of Sociology* 73:732-740.
Rosenbaum, James. 2001. *Beyond College for All*. New York: Russell Sage Foundation.
Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70:41-55.
—. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79:516-524.
Sabo, DF, BE Vanfossen, and MJ Melnick. 1993. "High School Athletic Participation and Post-Secondary Educational and Occupational Mobility: A Focus on Race and Gender." *Sociology of Sport Journal* 10:44-56.
Schafer, WE and M Armer. 1968. "Athletes are not Inferior Students." *Trans-Action* 6:61-62.
Schreiber, JB and EA Chambers. 2002. "After-School Pursuits, Ethnicity, and Achievement for 8th- and 10th- Grade Students." *Journal of Educational Research* 96:90-100.
Smith, Jeffrey and Petra Todd. 2005. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125:305-353.
Snyder, Eldon E. 1972. "High School Athletes and Their Coaches: Educational Plans and Advice " *Sociology of Education* 45:313-325.
Spady, WG. 1970. "Lament for the Letterman: Effects of Peer Status and Extracurricular Activities on Goals and Achievement." *American Journal of Sociology* 75:680-702.
Spreitzer, E and M Pugh. 1973. "Interscholastic Athletics and Educational Expectations." *Sociology of Education* 46:171-182.

APPENDIX A: Descriptive Statistics Tables

Table A1: Descriptive Statistics by Interscholastic Sports Participation

	Overall mean or %	Participants mean or %	Non-Participants mean or %	diff.	
GPA-12th grade	2.9571	3.0236	2.8639	0.1597	***
GPA-10th grade	2.8420	2.9261	2.7241	0.2020	***
interscholastic participation	58.37				
intramural participation	36.38	52.83	13.32	39.5100	***
participates in non-school sport	0.9875	1.1095	0.8164	0.2931	***
non-athletic participation	56.38	58.50	53.41	5.0900	***
Sex (boy ref.)					
Girl	51.80	48.69	56.17	-7.4800	***
Race (white ref.)					
Asian	9.68	7.42	12.85	-5.4300	***
black	10.81	10.46	11.30	-0.8400	
Hispanic	12.75	10.69	15.63	-4.9400	***
other	5.27	5.37	5.12	0.2500	
Family Structure (two-parent ref.)					
cohabiting parents	14.05	12.97	15.56	-2.5900	***
single parent	18.91	17.18	21.34	-4.1600	***
other household structure	3.06	2.46	3.91	-1.4500	***
SES	0.1119	0.2235	-0.0446	0.2681	***
mother's education	3.8632	4.0950	3.5383	0.5567	***
father's education	4.0233	4.2598	3.6916	0.5682	***
English as a second language	15.49	11.66	20.87	-9.2100	***
father unemployed	3.17	2.49	4.11	-1.6200	***
mother unemployed	7.44	6.13	9.28	-3.1500	***
academic track (academic ref.)					
college prep	59.26	64.05	52.55	11.5000	***
vocational	8.39	6.42	11.15	-4.7300	***
reading test score	31.0787	31.8475	30.0009	1.8466	***
math test score	47.5130	48.9991	45.4293	3.5698	***
educational expectations (HS diploma ref.)					
less than high school	0.33	0.28	0.39	-0.1100	
some college	7.97	6.28	10.34	-4.0600	***
four-year degree	37.01	37.04	36.97	0.0700	
greater than four-year degree	42.07	46.59	35.74	10.8500	***
not sure	8.72	7.09	11.00	-3.9100	***
School type (public ref.)					
Private	8.82	10.76	6.10	4.6600	***
Catholic	14.83	17.87	10.56	7.3100	***
urbanity (suburban ref.)					
urban	30.84	30.91	30.74	0.1700	
rural	19.03	19.34	18.58	0.7600	
N	9760	5697	4063		

Table A2: Descriptive Statistics by Popular Sports Participation

	Overall mean or %	Participants mean or %	Non-Participants mean or %	diff.	
GPA-12th grade	2.9571	2.9952	2.9492	0.0460	**
GPA-10th grade	2.8420	2.8631	2.8376	0.0255	
Popular Sports Participation	17.35				
intramural participation	36.38	53.04	32.89	20.1500	***
participates in non-school sport	0.9875	1.1607	0.9512	0.2095	***
non-athletic participation	56.38	55.05	56.66	-1.6100	
Sex (boy ref.)					
Girl	51.80	39.57	54.37	-14.8000	***
Race (white ref.)					
Asian	9.68	4.31	10.81	-6.5000	***
black	10.81	15.48	9.83	5.6500	***
Hispanic	12.75	8.45	13.65	-5.2000	***
other	5.27	5.61	5.19	0.4200	
Family Structure (two-parent ref.)					
cohabiting parents	14.05	13.82	14.09	-0.2700	
single parent	18.91	18.90	18.92	-0.0200	
other household structure	3.06	3.07	3.06	0.0100	
SES	0.1119	0.1519	0.1035	0.0484	**
mother's education	3.8632	3.9657	3.8417	0.1240	**
father's education	4.0233	4.0608	4.0154	0.0454	
English as a second language	15.49	7.56	17.16	-9.6000	***
father unemployed	3.17	2.89	3.22	-0.3300	
mother unemployed	7.44	6.14	7.71	-1.5700	*
academic track (academic ref.)					
college prep	59.26	61.25	58.84	2.4100	
vocational	8.39	6.62	8.76	-2.1400	**
reading test score	31.0787	30.5154	31.1970	-0.6816	**
math test score	47.5130	47.4336	47.5297	-0.0961	
educational expectations (HS diploma ref.)					
less than high school	0.33	0.47	0.30	0.1700	
some college	7.97	6.67	8.24	-1.5700	*
four-year degree	37.01	39.87	36.41	3.4600	**
greater than four-year degree	42.07	42.06	42.07	-0.0100	
not sure	8.72	8.09	8.85	-0.7600	
School type (public ref.)					
Private	8.82	14.35	7.66	6.6900	***
Catholic	14.83	14.53	14.89	-0.3600	
urbanity (suburban ref.)					
urban	30.84	27.58	31.52	-3.9400	***
rural	19.03	24.34	17.91	6.4300	***
N	9760	1693	8067		

APPENDIX B: Tables of Strata Frequencies and Mean Propensity Scores

Table B1: Interscholastic Sports:
Strata Frequencies and mean propensity scores

		Control			Treatment		
	Stratum	N	Mean	St Dev.	N	Mean	St Dev.
Low	1	509	0.1617	0.023	69	0.1623	0.024
	2	714	0.2508	0.029	235	0.2552	0.027
	3	338	0.3245	0.014	144	0.3257	0.015
	4	314	0.3746	0.014	186	0.3757	0.014
	5	611	0.4498	0.029	482	0.4514	0.029
	6	483	0.5475	0.029	599	0.5510	0.029
	7	199	0.6249	0.015	361	0.6255	0.014
	8	91	0.6617	0.007	178	0.6621	0.007
	9	51	0.6864	0.007	193	0.6875	0.007
	10	198	0.7477	0.028	773	0.7529	0.030
High	11	340	0.8678	0.039	2292	0.8760	0.037
		3848			5512		

Table B2: Whites in Interscholastic Sports:
Strata Frequencies and mean propensity scores

		Control			Treatment		
	Stratum	N	Mean	St Dev.	N	Mean	St Dev.
Low	1	12	0.1976	0.002	7	0.1959	0.001
	2	161	0.2248	0.014	26	0.2270	0.014
	3	192	0.2761	0.015	69	0.2800	0.014
	4	361	0.3496	0.029	165	0.3519	0.029
	5	450	0.4498	0.029	328	0.4537	0.028
	6	375	0.5518	0.028	474	0.5519	0.029
	7	227	0.6452	0.029	562	0.6497	0.028
	8	126	0.7437	0.029	535	0.7487	0.028
High	9	197	0.8929	0.043	1483	0.8918	0.043
		2101			3649		

Table B3: Blacks in Interscholastic Sports:
Strata Frequencies and mean propensity scores

	Stratum	Control N	Control Mean	Control St Dev.	Treatment N	Treatment Mean	Treatment St Dev.
Low	1	6	0.1912	0.004	3	0.1937	0.008
	2	128	0.2510	0.028	39	0.2638	0.026
	3	131	0.3501	0.028	71	0.3551	0.025
	4	101	0.4725	0.053	89	0.4878	0.055
	5	33	0.7365	0.059	117	0.7344	0.056
High	6	39	0.8684	0.037	256	0.8685	0.036
		438			575		

TableB4: Popular Sports:
Strata Frequencies and mean propensity scores

	Stratum	Control N	Control Mean	Control St Dev.	Treatment N	Treatment Mean	Treatment St Dev.
Low	1	374	0.0417	0.005	18	0.0397	0.004
	2	1916	0.0780	0.014	146	0.0807	0.012
	3	1032	0.1120	0.007	120	0.1126	0.008
	4	875	0.1372	0.007	154	0.1375	0.008
	5	1460	0.1729	0.014	303	0.1735	0.015
	6	873	0.2223	0.015	261	0.2238	0.014
	7	562	0.2737	0.014	221	0.2738	0.014
	8	585	0.3397	0.028	297	0.3395	0.027
High	9	92	0.4200	0.012	76	0.4218	0.013
		7769			1596		

Table B5: Whites in Popular Sports:
Strata Frequencies and mean propensity scores

	Stratum	Control N	Control Mean	Control St Dev.	Treatment N	Treatment Mean	Treatment St Dev.
Low	1	737	0.0858	0.010	70	0.0858	0.010
	2	734	0.1121	0.007	78	0.1122	0.006
	3	616	0.1376	0.007	102	0.1383	0.007
	4	1106	0.1738	0.014	244	0.1742	0.015
	5	702	0.2227	0.014	198	0.2242	0.014
	6	386	0.2747	0.014	174	0.2745	0.015
	7	375	0.3395	0.027	180	0.3422	0.028
High	8	23	0.4063	0.003	21	0.4058	0.004
		4679			1067		

Table B6: Blacks in Popular Sports:
Strata Frequencies and mean propensity scores

		Control			Treatment		
	Stratum	N	Mean	St Dev.	N	Mean	St Dev.
Low	1	172	0.0816	0.011	12	0.0763	0.012
	2	230	0.1421	0.029	45	0.1484	0.032
	3	221	0.2818	0.058	94	0.2982	0.053
	4	67	0.4480	0.029	47	0.4549	0.030
	5	29	0.5391	0.028	38	0.5471	0.030
High	6	9	0.6287	0.015	9	0.6273	0.017
	Total	728			245		