High School Clubs Participation and Earnings

Vasilios D. Kosteas
Cleveland State University
2121 Euclid Avenue, RT 1707
Cleveland, OH 44115-2214
b.kosteas@csuohio.edu
Tel: 216-687-4526, fax: 216-687-9206

Abstract

This paper estimates the effect of participation in high school extracurricular activities on future earnings, making three important contributions to the existing literature: 1) it compares the earnings effects of participation in different types of clubs; 2) it investigates whether the effect of clubs participation is constant over time; and 3) it employs an instrumental variables approach in order to identify a causal link between clubs participation and wages. Using the NLSY79 dataset, I find that participation in high school clubs leads to higher future earnings. While previous studies have focused on athletics, I show that participation in both athletics and academic clubs have positive earnings effects. These results are robust to various estimation routines and robustness checks.

JEL codes: J31

Keywords: earnings, high school clubs, NLSY

Introduction

One of the debates in school funding revolves around the number and types of extracurricular activities offered to students. Many school districts respond to financial shortfalls by cutting back on these programs and focusing on "core" educational programs, at the expense of the arts, sports and other activities. Opposition to these funding cuts rests on the belief that these activities are an important part of students' educational and personal growth. Supporters of these clubs and activities believe they help build skills, such as the ability to communicate and work with others, which are valuable to future employers. This paper addresses this issue by focusing on the link between participation in different types of clubs during high school and future wages.

This paper makes three important contributions to the existing literature: 1) it compares the earnings effects of participation in different types of clubs; 2) it investigates whether the effect of clubs participation is constant over time; and 3) it employs an instrumental variables approach in order to identify a causal link between clubs participation and wages. Previous studies have shown a positive link between participation in high school athletics and future wages. This paper seeks to determine whether this effect is limited to athletics or if it extends to other types of clubs, particularly extracurricular activities with an academic or leadership focus. Additionally, most studies in this literature use a single year's worth of labor market data. This limits each study's focus to a point in time, not allowing for an investigation into how the labor market effects of participation in high school extracurricular activities might vary over the life-cycle. Finally, most papers have attempted to address the causality issue by using various instrumental variables. However, recent results in other studies call into question the validity of these instruments.

Literature review and motivation

Participation in extracurricular activities can lead to improved labor market outcomes if it results in increased human capital for the involved students. It can lead to greater opportunities for moving up the firm hierarchy, resulting in more frequent (and possibly bigger) promotions and leading to high-level supervisory positions if participation builds leadership, organizational or social skills. Time constrained students must choose how to allocate their time between their formal studies and activities outside of the classroom. Participating in extracurricular activities can lead to greater human capital acquisition if it is an investment good, but may lower human capital if it takes time away from the student's studies. The latter effect depends on how students allocate their time and substitute between activities. Existing research shows that student-athletes spend more time studying than non-athletes and female student athletes spend less time watching television (Anderson, 2001). Even in cases when involvement in clubs and sports teams detracts from time spent studying, the net effect on human capital acquisition still may be positive if the student builds more human capital through clubs participation than she loses from the reduced time spent on her academics. Furthermore, the skills developed through active participation in extracurricular activities may be different than those which are developed in the classroom. Thus, sacrificing some classroom performance in favor of club activities may be the optimal strategy for maximizing future labor market outcomes, particularly when there are diminishing returns to specific skills with respect to earnings.[1]

It is also possible that active involvement in clubs can enhance student learning in the classroom. Using a fixed effects model, Lipscomb (2007) finds that participation in athletics is positively correlated with both science and math test scores while getting involved in clubs

[1] It may also be the case that the skills acquired during extracurricular activities are complimentary to those built in the classroom, raising the return to the latter.

results in higher math test scores. Lipscomb's findings indicate that athletic participation reinforces classroom performance, complicating the relationship between clubs participation and human capital acquisition even further. In the absence of direct and complete measures of human capital, researchers must examine the link between clubs participation and educational and labor market outcomes to determine whether these activities provide non-consumption benefits to participants.

The existing empirical literature in economics has focused mainly on the effect of participation in high school athletics on educational and labor market outcomes. Barron, Ewing and Waddell (2000) analyze the relationship between participation in high school athletics on high school rank, educational attainment, future employment and weekly wages. They find no significant employment effect of participation in high school athletics; however, men who participated in high school athletics earn significantly higher wages than those who did not. The authors use two data sets, the National Longitudinal Study of the High School Class of 1972 and the National Longitudinal Survey of Youth. The latter is the same dataset used in the present paper; however, the authors only examined labor market outcomes for the year 1992 in the NLSY dataset. Furthermore, their focus on the intensity of involvement in athletics leads to a dramatically reduced sample size (891 observations for the earnings equations and 1,047 observations for the employment equations) since information on the extent of individual's involvement was not available for all respondents.

Eide and Ronan (2001) estimate the relationship between athletics participation in high school and educational attainment and future wages. They estimate models for dropping out of high school, attending college, finishing college, and annual earnings. The authors estimate each model using a single cross-section (different cross-sections are used to investigate the different dependant variables). Their results are mixed, finding a negative correlation with educational attainment for white men, a positive correlation for black men and white women, and no correlation for Hispanics or black women. They do not find a significant correlation between athletics and wages for any of these groups. Using the 1990 wave of the NLSY79, Ewing (2007) finds that in addition to earning higher wages, former high school athletes also receive more benefits. Anderson (2001) finds a positive link between participation in sports and educational outcomes for white athletes. There are also several empirical papers in the sociology literature which investigate the link between high school athletics and various educational outcomes (see Anderson, 2001 for a summary of those papers).

Kuhn and Weinberger (2005) focus on the effect of leadership positions on labor market outcomes. They find that men who hold leadership positions in high school are more likely to be in managerial occupations and earn higher wages. The focus on leadership positions is interesting and important; however, few students ever hold these positions, so their results cannot speak to the broader impact of participation in high school extracurricular activities and labor market outcomes. Kuhn and Weinberger use multiple datasets from different decades. By using different datasets, the authors can attempt to see whether the returns to leadership skills change over time. However, the authors use labor market data from follow up interviews which occurred between nine and thirteen years after the initial interview. Thus, they are unable to comment on how the earnings effect of leadership skills measured during high school varies over the life cycle.

This review of the empirical literature shows that studies have generally focused on a single type of high school activity (athletics or leadership positions) and generally use a single year's worth of data to analyze a particular labor market outcome. Thus, these studies are unable either to compare the effects of participation in different types of extracurricular activities or to examine whether these effects vary over the life cycle. This paper contributes to the literature by focusing on those two questions. Additionally, I introduce a new approach to account for the potential endogeneity of the clubs participation variables, which allows me to speak to whether participation leads to higher future income or simply whether people with greater income earning potential are also more likely to participate in extracurricular activities.

Data and estimation routine

Estimation strategy

This paper's primary exercise is to estimate the effect of participation in high school extracurricular activities on future income. The simple approach estimates these effects by fitting the model via ordinary least squares. However, there are two potential problems with this approach: 1) endogeneity of clubs participation, and 2) measurement error in the clubs variable. The first econometric issue has been well documented in the existing literature and previous studies have addressed it to varying degrees. Students who participate in clubs also may possess other attributes/skills which are valued in the labor market, leading to an upward bias in the estimated effect of clubs participation on labor market outcomes. On the other hand, there may be a negative selection into certain types of extracurricular activities; the most academically gifted students may be more inclined to participate in academic clubs but avoid athletics and other nonacademic activities. Thus, it is difficult to predict the direction of the endogeneity bias ex ante. The simple way to address this issue is through the use of individual fixed effects. However, the key variable in the model (clubs participation in high school) is only observed once, ruling out fixed effects as a possible candidate.

The second problem, measurement error, will lead to a downward bias in the estimated effect of clubs participation on labor market outcomes. Measurement error is a potentially significant problem in this dataset for multiple reasons. First, information on clubs participation is collected in 1984, when respondents are between the ages of nineteen and twenty-seven. Thus, the survey is asking people to report on activities that occurred up to ten years earlier. Second, some individuals may report that they were involved in a particular type of activity, even if their participation was short-lived or superficial. With both endogeneity and measurement error, the direction of the bias cannot be determined ex ante.

Other papers in the literature have taken various approaches in dealing with the potential endogeneity of the participation variable. As a robustness check, Barron et al (2000) instrument for participation in athletics with school size and other socioeconomic and school quality variables. However, both previous studies (Betts, 1995; Hanushek, 1986) and results in this paper show that school characteristics affect student and labor market outcomes. When analyzing the effects of holding leadership positions in high school on labor market outcomes, Kuhn and Weinberger (2005) control for family background and high school fixed effects. Robustness checks which add controls for psychological traits as well as beauty and height support their main results. While their series of robustness checks provides some evidence that their findings may capture a true correlation, they do not control for individual fixed effects or employ an IV estimation routine which can account for the potential correlation between unobserved ability and leadership roles. Eide and Ronan (2001) use height as an instrument for high school sports participation. I do not take this approach in the current paper for two reasons. First, while height shows a strong correlation with participation in athletics, it is not highly correlated with general extracurricular participation. Second, it is not clear that height is a valid instrumental variable; other studies have shown that height is correlated with labor market outcomes (Persico et al 2004).

This paper uses two alternative approaches to deal with the potential endogeneity of the clubs variable. The first approach controls for household fixed effects by transforming all variables into deviations from household means. This will account for any potentially confounding effects of the individual's family background, such as parental education or socioeconomic status. The second approach employs instrumental variables estimation using information on siblings' clubs participation to instrument for the clubs participation variable. Given the finding by Rees, Lopez, Avarett and Argys (2008) that birth order is correlated with participation in athletics and other extracurricular activities, I create an indicator variable for whether the individual is the firstborn sibling as an additional instrument.

The use of sibling data has been employed in various studies, most notably in Ashenfelter and Krueger (1994), which uses data on twins to estimate returns to schooling. By using data on multiple individuals from the same household, the authors are able to control for family specific fixed effects by transforming the data into deviations from household means. Then, to account for potential measurement error, the authors use the sibling's report of the other respondent's educational attainment to instrument for the respondent's self reported level of schooling. This approach has been used in several papers to estimate the returns to schooling, including Ashenfelter and Zimmerman (1997) which uses data on brothers and fathers and sons, and Bronars and Oetinger (2006) which uses sibling data from the NLSY79, amongst others.   The Bronars and Oetinger paper is particularly relevant to the present study since it employs the same dataset.

Consider the following specification for the empirical model:

$$y_{ijt} = \beta x_{ijt} + \nu_i + \upsilon_j + \varepsilon_{ijt}\,, \qquad (1)$$

where y is the log of average weekly earnings for individual i in household j at time t, x is a vector of individual and firm characteristics, ν_i is an individual fixed effect, υ_j is a family (or household) fixed effect and ε_{ijt} is an iid disturbance. It is assumed that both the individual and the household fixed effects are correlated with unobserved factors which fall into the error term. Thus, fitting the model using least squares estimation will yield biased estimates. The household fixed effect captures all aspects of the home environment which are common to each individual in that household and may reflect both environmental (nature) and genetic (nurture) commonalities. Transforming each variable in the model into deviations from household means eliminates the household fixed effect, but does not eliminate the individual-specific fixed effect,

$$dy_{ijt} = \beta dx_{ijt} + d\nu_i + d\varepsilon_{ijt}\,. \qquad (2)$$

Thus, this transformation of the data can eliminate part of the potential bias due to endogeneity; however there is still the potential bias due to a correlation between the individual fixed-effect and the error term.

To account for the potential endogeneity due to unobserved individual characteristics, I use the average of the siblings' clubs participation as an instrument for each respondent's clubs participation in high school. This approach will also eliminate the downward bias from measurement error if the measurement errors for own clubs participation and siblings' participation are not systematically correlated. Siblings' clubs participation is likely to be correlated with household effects; however this will not pose a problem if the household effects do not exert any influence on the coefficient estimates. Additionally, we can attempt to proxy for some of the factors that might affect the fixed effect. Specifically, siblings' clubs participation is likely to be correlated with parental education and high school size and quality.

As a robustness check, variables for mother's and father's education, school size and quality (measured by the percent of students who are disadvantaged and the percent of students who drop out by the tenth grade) are included as explanatory variables in an expanded model. Tests are provided to support the exclusion restriction for the instruments for each set of instrumental variables estimates. To facilitate these tests and strengthen the first stage estimation, an indicator variable for whether the individual is the first born child in the household is included as a second instrument. Including a second instrument allows estimation of Hansen's J-statistic which tests whether the instruments have been properly excluded from the second stage regression.

Bronars and Oetinger (2006) use sibling data from the NLSY79 to instrument for self-reported schooling. In this dataset, there are many individuals who have more than one sibling in the survey. For these individuals, they took the sibling report which was determined to be the most likely to be accurate. However, this throws out potentially useful information, so I use the average value of the clubs variables across all siblings to construct the instrumental variables. The instrumental variables are constructed using the information on clubs participation for all siblings in the dataset, regardless of whether or not they are included in the estimation sample.

3.2 Data and variables

The empirical estimation conducted in this paper makes use of the 1984-2006 waves of the NLSY79. The NLSY, which conducted surveys every year starting in 1979 through 1994, then in even numbered years, began with an initial sample of 12,686 individuals. The initial sample contained oversamples of poor white individuals and members of the armed forces. The military and poor white oversamples were dropped in 1985 and 1991, respectively. Attrition also leads to further loss of sample size; while 9,018 individuals were surveyed in 1991, this number falls to 7,654 in the 2006 wave of the survey. I begin the estimation sample using the 1984 survey for multiple reasons. First, information on clubs participation comes from the 1984 wave of the survey. Second, at this point the youngest respondents in the survey are nineteen years old, which should put all or nearly all individuals in the sample past high school age. The twenty-seven year span of the sample covers individuals from nineteen to fifty years of age, enabling me to analyze how the earnings effects of clubs participation might vary over the life cycle. All nominal dollar values are converted to 2006 dollars using the CPI.

The empirical model consists of an augmented Mincerian wage equation with controls for club participation, a firm size variable, whether the individual is covered by an employee association contract, and individual characteristics. Each model is estimated using three estimators: ordinary least squares, household fixed effects (HFE) and instrumental variables (IV) estimated via two-stage least squares. Then, robustness checks are performed by including school quality variables and parental education as additional explanatory variables. These variables are excluded in the basic estimation due to availability; their use results in a substantial additional loss of observations.

The NLSY data set contains information on the types of clubs the respondent was a member of in high school, with eight categories. These categories are: youth organizations, hobby, student government, yearbook and newspapers, athletics, performance art, honor society and other clubs. The information on high school clubs participation was collected in 1984, when the respondents were between the ages of nineteen and twenty-seven. Borghans, ter Wheel and Weinberg (2006) use this information to construct a variable which sums up the number of types of clubs in which the individual participated. I take the same approach here. Thus, the clubs variable takes integer values between zero (did not belong to any clubs) to eight (participated in clubs in all eight categories).

The clubs variable does not capture the total number of clubs in which an individual was active, or the extent of the individual's involvement. An individual only participating in athletics may have been involved with more groups (playing three different sports) than someone who participated in two clubs in different categories. While the clubs variable is not a perfect measure of participation in extracurricular organizations, it does contain useful information on social activities. Using the clubs variable as a measure of social skills, Borghans et al (2006) find that sociability has an impact on wages and choice of occupation. It is also likely that these skills are very important in management positions, affecting a person's ability to supervise others and making these skills an important determinant of promotions and the assignment of supervisory responsibility.

In addition to the basic clubs variable, a dummy variable for clubs participation, taking a value equal to one if the individual participated in any type of clubs and zero otherwise is constructed and used to perform robustness checks. I expect the results to be qualitatively the same as those obtained using the primary clubs participation variable; however the magnitude of the clubs participation effect may differ. This may arise if there are diminishing returns to participating in multiple types of clubs. Participating in any clubs at all may be more important than being involved in many clubs. In fact, given time constraints, participating in too many clubs may lead to lower future earnings by leading to lower educational outcomes today.

In order to test whether participation in different types of clubs has the same impact on future earnings, I also construct indicator variables for participation in the following three categories of extracurricular activities: athletics, academic (honor society, yearbook and student government) and nonacademic clubs (youth organizations, hobby, performing arts and other types of clubs). Separating out athletics as an independent category allows for a comparison of the results obtained in this paper with previous studies. Separating the remaining clubs into academic and nonacademic simply seems to provide a logical grouping.

In addition to the clubs variable, the model includes several individual controls including gender, race (black and Hispanic dummies), age, age squared, log of tenure in years and its square, years of schooling completed, AFQT score percentile, the average number of hours worked per week and whether the respondent is covered by a union or employee association contract. Extended models also include the log number of employees at the respondent's location as a measure of firm size. All models include year as well as industry and occupation indicators at a moderate degree of aggregation; there are thirty industry and twenty-four occupation categories. The dependant variable is average weekly earnings, which is constructed by dividing earnings over the past calendar year by the number of weeks employed.

Observations are excluded from the sample for the following reasons: if any of the key variables are missing, the individual was employed in the agricultural sector or in the active armed forces, if the respondent worked fewer than five-hundred or more than thirty-five hundred hours in the past calendar year or had an average weekly income less than one-hundred dollars. The hours worked restrictions are intended to exclude seasonal workers and individuals who work relatively few hours per week. This leaves a full estimation sample consisting of 77,980 observations 9,498 individuals. The HFE sample contains 44,806 observations while the IV sample contains 43,398 observations. Both of these samples only use observations where observations for at least one sibling are available.

Before turning to the results, I provide a thorough description of the sample data. Table one provides the frequency of the clubs participation variable for both the full and siblings samples. Focusing on the siblings sample, we see that over thirty-seven percent of individual-year observations are for respondents who did not participate in any high school extracurricular activities. Nearly one-third participated in only one type of club while slightly less than one-third participated in two or more types of clubs. A very small percentage participated in five or more types of clubs. Columns five and six provide the frequencies for the clubs variable using only data from the 1984 survey. The statistics match those for the siblings sample fairly closely, indicating that the estimation sample is fairly balanced in the following sense: there is no bias towards individuals with either high or low clubs participation in having a larger average number of observations. Finally, comparing the siblings sample to the full sample, we can see that the more restricted sample is broadly representative of the full sample, at least in terms of clubs participation during high school.

Table two provides summary statistics for the full (columns 1-2) and siblings samples (columns 3-4). A quick look at the two sets of summary statistics supports the conclusion from table one that the more restricted sample is representative of the larger sample. Focusing on the latter, we find that the average individual participated in roughly one and one-fourth types of clubs, while roughly over sixty percent of respondents participated in some type of club. Over forty percent report having participated in athletics while nearly thirty percent participated in academic clubs and thirty-six percent participated in nonacademic type clubs.

4. Results

The results for the model using the clubs participation variable are presented in table three. Column one provides the coefficient estimates obtained when the model is fitted via ordinary least squares using the full sample. The results show a positive link between participating in clubs and weekly earnings; participating in one additional type of club is associated with a nearly one percent increase in weekly earnings. This effect is equal to roughly two-ninths of the wage effect of an additional year of school or a three point increase in the AFQT score percentile (which is approximately one-ninth of a standard deviation for that variable). All other variables have the expected signs; wages increase with age (at a decreasing rate), job tenure, and when the individual is covered by a union contract. Being a woman or black are both associated with lower wages, while being Hispanic is associated with higher wages. The model is also fitted via OLS but on the sample restricted to individuals who have a sibling in the survey to check for the consistency of the results across the two samples; this is done to rule out the possibility that differences in the results obtained via HFE or IV estimation are not being driven by the change in the estimation sample. The results (not presented in the table but available from the author upon request) are highly similar to those presented in column one.

Next, we turn our attention to the model estimated after transforming the variables into deviations from household means (column two). This approach should eliminate the fixed household component which may be correlated with both clubs participation during high school and future earnings. The results continue to show a positive, highly statistically significant correlation between clubs participation and earnings; however the magnitude has increased significantly. Participating in one additional type of club is now associated with a 2.1 percentage point increase in weekly earnings, compared to the one percent increase found when employing OLS estimation. The estimated effect is now equal to nearly half an additional year of schooling or a nearly eight point increase in the individual's AFQT score percentile.

Finally, column three presents the coefficient estimates when the model is fitted via two-stage least squares. The effect of clubs participation is even greater than the effect found using HFE estimation. Participation in an additional type of club yields a future earnings gain of 2.8 percent, which is more than the increase from an additional half year of schooling. The impact is now estimated as equivalent to a ten point increase in the AFQT score percentile (which is more than one-third of a standard deviation). The partial R-squared of the instrumental variables from the first stage equation along with the Anderson statistic both indicate that the instruments are not weak.[2] The Hansen statistic fails to reject the validity of the instruments, indicating that they have been correctly excluded from the second stage equation as separate explanatory variables. The increase in the coefficient on the clubs variable over the results presented in column one indicates that the downward bias due to measurement error outweighs the endogeneity bias when fitting the model vial OLS.

As a robustness check for the specification of the clubs variable, the basic model is also fitted using each of the three estimators including the clubs participation indicator variable in place of the number of types of clubs variable. Since the coefficient estimates for the other variables are nearly identical to those in shown it table three, table four provides only the results for the clubs indicator variable. The results reinforce those found in table three; participating in high school clubs has a significant effect on future earnings. Focusing on the IV estimates, however we now see an even larger impact; participating in extracurricular activities raises

[2] We must always bear in mind that these tests cannot definitively rule out the possibility that the instruments are weak. However, the Anderson statistic is well above the standard levels that are accepted as supporting the use of the instruments.

future earnings by 11.8 percent. This is the same effect on earnings as more than two-and-a-half additional years of schooling. Overall, the results provided in tables three and four show that high school clubs participation has a positive and significant (both economically and statistically speaking) impact on future earnings.

Effects by club type

Next, I look at the effects of participating in different types of clubs. For this model, a separate instrument is constructed for each of the three endogenous variables when employing IV estimation, in addition to the first born indicator variable. Once again, the test statistics support the validity of the instruments. As noted in the literature review, most of the existing literature has focused on the labor market effects of participation in high school athletics, finding mixed results, particularly with respect to wages. Examining three categories of activity: athletics, academic and nonacademic clubs (excluding athletics), I find significant differences between different types of clubs (table five). Estimates using each of the three estimators show a negative and statistically significant correlation between participation in nonacademic clubs and future earnings and a positive effect of participation in high school athletics on earnings. Based on the IV results, it appears that athletics participation leads on average to a nineteen percent increase in future weekly earnings. The results are consistent with Barron, Ewing and Waddell (2000) who found a positive wage effect of participation in high school athletics. Both the OLS and HFE estimates show a positive effect of participation in academic clubs on future earnings; however the estimate turns negative (but not statistically significant) when using 2SLS estimation. The IV statistics indicate that we should interpret these results with caution. The Hansen statistic shows that we can reject the null hypothesis that the instruments are endogenous at the twelve percent level. Furthermore, the first stage R-squares are very low; a fact which is also reflected in the substantial increase in the standard errors from the second stage. Overall, the present results show that the positive earnings effect of clubs participation is not limited to athletics.

Effects over the life cycle

To test whether the effects found in table three are steady over the life-cycle, I fit the model including an interaction term between clubs participation and the individual's age. To account for the additional endogenous explanatory variable, the IV estimation routine employs the sum instruments as the basic model: siblings' average clubs participation and an indicator variable for whether the individual is firstborn. The results, provided in table six, show an initially negative effect of clubs participation which becomes positive over time. According to the HFE estimates (column two) the effect becomes positive at twenty-three years of age. Thus, the effect is positive for almost the entire age range of our sample (which is nineteen to fifty years old). Interestingly, the overtaking age occurs at about the age when most college graduates would have begun their careers. This raises the question of whether there may be different effects of clubs participation for college graduates relative to those who never attended college. The IV estimates show the same pattern; however the coefficients are not statistically significant. Overall, the results indicate that the effect of clubs participation on earnings grows stronger over time. One potential explanation for this result is that the skills developed while participating in these activities interact with or increase the rate at which individuals build human capital through post-school training.

Robustness check: parents' education and school quality variables

As an additional robustness check, the model is estimated via two-stage least squares including the mother and father's highest grade completed and school quality variables as additional controls. I include the log number of students at the respondent's school at age fourteen and two school quality measures: the average daily attendance rate and the percent of students who have dropped out by the tenth grade. The means (standard deviations) for these variables are 88.48 (15.96) and 15.48 (20.16), respectively. Including these two variables leads to a further reduction in sample size; the siblings sample now contains 22,517 observations. The results, presented in table seven, continue to show a positive, significant relationship between participating in clubs and future earnings. In fact, the estimated effects are somewhat larger; participating in one additional type of club is now found to increase weekly earnings by 3.5 percent compared with 2.8 percent (from table three). The model which separates clubs participation into three categories continues to show that nonacademic clubs participation does not have any effect on future earnings while athletics do confer earnings gains. We now see no significant correlation between participation in academic clubs and earnings. However, the standard errors on the coefficient estimates are so large, that the differences between the coefficients are not statistically significant at conventional levels.

Conclusions

This paper estimates the effect of participating in high school extracurricular activities on future earnings. I find that clubs do indeed have a significant, positive impact on future earnings. The effect is roughly equal to one-half of an additional year of formal schooling. The results are robust to various estimation techniques and robustness checks which account for the potential endogeneity of the clubs participation variable. By using instrumental variables estimation, this paper advances the literature by showing that the relationship between clubs participation and earnings goes beyond mere correlation, supporting the position that there is a causal link between the two. Additionally, where the existing literature has mostly focused on the labor market effects of participation in high school athletics, this paper shows that academic clubs may also have a significant, positive effect on future earnings. However, the results by club type are mixed and should be interpreted with caution.

These findings have serious implications for debates over school funding. They show that participating in extracurricular activities yields benefits beyond consumption value. The higher future earnings of participants suggest that there is a labor market value to these activities, but these findings alone do not indicate whether the investment is a wise one. That calculation requires a look at all of the costs and benefits- higher earnings, employment rates, and of course the consumption value of these activities, both for the participants and the community. The lack of a positive earnings effect associated with nonacademic clubs should not be taken as evidence against their provision. After all, performing in the school play may provide critical preparation for and ultimately access to a career in the performing arts. The exercise undertaken in this paper merely sheds light one aspect of the potential benefits obtained from participating in extracurricular activities during high school.

References

Anderson, D. J. (2001). 'If You Let Me Play': The Effects of Participation in High School Athletics on Students' Educational and Labor Market Success. Mimeo.

Ashenfelter, O. & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review,* 84(5), 1157-1173.

Ashenfelter, O. & Zimmerman, D. J. (1997). Estimates of the returns to schooling from sibling data: fathers, sons and brothers. *The Review of Economics and Statistics*, 79(1), 1-9.

Barron, J. M., Ewing, B. T., & Waddell, G. R. (2000). The effects of high school athletics participation on education and labor market outcomes. *The Review of Economics and Statistics,* 82, 409-421.

Borghans, L., ter Wheel, B., & Weinberg, B. A. (2006). People people: social capital and the labor-market outcomes of underrepresented groups. NBER working paper #11985.

Borghans, L., ter Wheel, B., & Weinberg, B. A. (2008). Interpersonal Styles and Labor Market Outcomes. *Journal of Human Resources,* 63(4), 816-858.

Bronars, S. G. & Oettinger, G. S. (2006). Estimates of the return to schooling and ability: evidence from sibling data. *Labour Economics,* 13, 19-34.

Eide, E. R., & Ronan, N. (2001). Is participation in high school athletics an investment or a consumption good? *Economics of Education Review,* 20, 431-442.

Ewing, B. T. (2007). The labor market effects of high school athletic participation. *Journal of Sports Economics,* 8, 255-265.

Kuhn, P. & Weinberger, C. (2005). Leadership Skills and Wages. *Journal of Labor Economics,* 23(3), 395-436.

Lipscomb, S. L. (2007). "Secondary school extracurricular involvement and academic achievement: a fixed effects approach." *Economics of Education Review,* Vol. 26, 463-472.

Persico, N., Postlewaite, A., & Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: the case of height. *Journal of Political Economy*, 112, 1019-1053.

Rees, D.I., Lopez, E., Averett, S. L., & Argys, L. M. (2007). Birth order and participation in school sports and other extracurricular activities. *Economics of Education Review,* 27, 354-362.

Table 1: Frequency distribution for clubs participation

	Full Sample		Siblings Sample		Siblings 1984	
Number of types of clubs:	Observations	Percent	Observations	Percent	Observations	Percent
0	29,716	38.11	17,381	38.79	1200	39.04
1	23,310	29.89	13,732	30.65	906	29.47
2	12,445	15.96	6,486	14.48	485	15.78
3	6,727	8.63	3,787	8.45	245	7.97
4	3,779	4.85	2,199	4.91	154	5.01
5	1,564	2.01	915	2.04	57	1.85
6	426	0.55	303	0.68	26	0.85
7	13	0.02	3	0.01	1	0.03
Total	77,980	100	44,806	100	3,074	100

Table 2: Summary Statistics

	Full Sample		Siblings Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Weekly income	660.97	531.47	662.2	533.17
First born	0.222	0.416	0.189	0.391
Clubs	1.205	1.32	1.189	1.332
Clubs indicator	0.619	0.486	0.612	0.487
Athletic clubs indicator	0.422	0.494	0.421	0.494
Academic clubs indicator	0.293	0.455	0.286	0.452
Non-academic clubs indicator	0.359	0.48	0.351	0.477
Hours worked per week	41.78	8.92	41.8	8.88
Highest grade completed	13.09	2.311	13.11	2.308
AFQT score percentile	46.36	28.57	42.1	28.65
Age	30.6	6.56	30.42	6.522
Tenure in years	3.912	4.271	3.99	4.339
Union contract indicator	0.139	0.346	0.142	0.349
Female	0.485	0.5	0.467	0.499
Black	0.248	0.432	0.26	0.439
Hispanic	0.166	0.372	0.161	0.367
Observations	77,890		44,806	

Columns 1-2 provide the mean and standard deviations of the key variables for the full sample.

Columns 3-4 provide the mean and standard deviations of the key variables for the siblings sample.

Table 3: Effect of clubs participation on earnings

	OLS	HFE	IV
Clubs participation	0.01**	0.021**	0.028**
	(0.0015)	(0.003)	(0.008)
Hours worked per week	0.02**	0.019**	0.019**
	(0.00025)	(0.00034)	(0.00035)
Highest grade completed	0.045**	0.044**	0.045**
	(0.0011)	(0.0021)	(0.0022)
AFQT score percentile	0.0031**	0.0027**	0.0028**
	(0.00009)	(0.00018)	(0.00016)
Age	0.087**	0.105**	0.081**
	(0.0039)	(0.005)	(0.0056)
Age squared	-0.0012**	-0.0014**	-0.0011**
	(0.000064)	(0.000078)	(0.000091)
Log of job tenure in years	0.07**	0.063**	0.071**
	(0.0015)	(0.0019)	(0.002)
Log of job tenure in years squared	0.019**	0.0167**	0.019**
	(0.00074)	(0.00092)	(0.001)
Union contract indicator	0.168**	0.109**	0.161**
	(0.005)	(0.0067)	(0.0066)
Female	-0.205**	-0.227**	-0.197**
	(0.0041)	(0.0067)	(0.0058)
Black	-0.048**	0.121	-0.068**
	(0.0047)	(0.116)	(0.0064)
Hispanic	0.054**	-0.142**	0.046**
	(0.0049)	(0.066)	(0.0066)
R-squared	0.49	0.3909	0.4913
Observations	77,980	44,806	43,398
Instrumental Variables Tests:	1st Stage Partial R-squared:		0.0611
	Anderson statistic (p-value):		2733.8 (.00)
	Hansen J-statistic (p-value):		0.54 (.46)

Instruments: average of clubs variable for siblings and an indicator variable for whether the individual is the firstborn child.

All models include year, industry and occupation indicator variables.
Robust standard errors are in parentheses.

The null hypothesis for the Hansen statistic is the instruments are exogenous and are correctly excluded from the first stage.

Table 4: Effect of clubs participation on earnings

	OLS	HFE	IV
Clubs participation dummy variable	0.018**	0.034**	0.118**
	(0.0038)	(0.0071)	(0.032)
R-squared	0.4899	0.3904	0.4876
Observations	77,980	44,806	43,398
Instrumental Variables Tests:	1st Stage Partial R-squared:		0.0247
	Anderson statistic (p-value):		1087.1 (.00)
	Hansen J-statistic (p-value):		0.681 (.41)

Instruments: average of clubs indicator variable for siblings and an indicator variable for whether the individual is the firstborn child.

All models include year, industry and occupation indicator variables.

Robust standard errors are in parentheses.

The null hypothesis for the Hansen statistic is the instruments are exogenous and are correctly excluded from the first stage.

Table 5: Effect of clubs participation on earnings by club type

	OLS	HFE	IV
Athletics	0.034**	0.029**	0.19**
	(0.0036)	(0.0068)	(0.042)
Academic clubs	0.022**	0.045**	-0.018
	(0.0042)	(0.0078)	(0.062)
Nonacademic clubs	-0.025**	-0.016**	-0.082†
	(0.0038)	(0.0069)	(0.047)
R-squared	0.4907	0.3909	0.4779
Observations	77,980	44,806	43,398

Instrumental Variables Tests:	1st Stage Partial R-squared:	
	Athletics	0.0367
	Academics	0.0195
	Non-academic	0.026
	Anderson statistic (p-value):	203.5 (.00)
	Hansen J-statistic (p-value):	2.413 (.12)

Instruments: average of clubs indicator variables for each type of club for the respondent's siblings and an indicator variable for whether the individual is the firstborn child.

All models include year, industry and occupation indicator variables.

Robust standard errors are in parentheses.

The null hypothesis for the Hansen statistic is the instruments are exogenous and are correctly excluded from the first stage.

Table 6: Effect of clubs participation on earnings over the life-cycle

	OLS	HFE	IV
Clubs	-0.082**	-0.069**	-0.257
	(0.0066)	(0.0086)	(0.391)
Age	0.084**	0.103**	0.076**
	(0.0039)	(0.005)	(0.0092)
Clubs*Age	0.003**	0.003**	0.0093
	(0.00021)	(0.00027)	(0.013)
R-squared	0.4916	0.3931	0.4872
Observations	77,980	44,806	43,398
Instrumental Variables Tests:	1st Stage Partial R-squared:		
		Clubs	0.0611
		Clubs*Age	0.0575
	Anderson statistic (p-value):		18.14 (.00)

Instruments: average of clubs variable for siblings and an indicator variable for whether the individual is the firstborn child.

All models include year, industry and occupation indicator variables.

Robust standard errors are in parentheses.

The null hypothesis for the Hansen statistic is the instruments are exogenous and are correctly excluded from the first stage.

Table 7: Robustness check

	1	2	3
Clubs	0.035**		
	(0.012)		
Clubs indicator variable		0.028	
		(0.047)	
Athletics			0.105†
			(0.064)
Academic clubs			-0.0049
			(0.087)
Nonacademic clubs			-0.016
			(0.077)
Mother's highest grade completed	0.0038**	0.0046**	0.0043**
	(0.0015)	(0.0015)	(0.0015)
Father's highest grade completed	0.0056**	0.0057**	0.0049**
	(0.0011)	(0.0012)	(0.0016)
Log school size	0.059**	0.048**	0.057**
	(0.0067)	(0.0063)	(0.0076)
Percent students disadvantaged	-0.00083**	-0.00073**	-0.00068**
	(0.00016)	(0.00016)	(0.00024)
Percent students drop out by	0.00064**	0.00062**	0.00061**
the tenth grade	(0.00016)	(0.00016)	(0.00016)
R-squared	0.5131	0.5133	0.5128
Observations	22,517	22,517	22,517
Instrumental variables tests:			
Anderson statistic (p-value):	1124 (.00)	510.3 (.00)	115.6 (.00)
Hansen J-statistic (p-value):	0.058 (.81)	0.398 (.53)	0.68 (.44)

All models include year, industry and occupation indicator variables.
Robust standard errors are in parentheses.
The null hypothesis for the Hansen statistic is the instruments are exogenous and are correctly excluded from the first stage.