THE IMPACT OF PARTICIPATION IN COMMUNITY AND SCHOOL BASED ACTIVITIES ON EDUCATIONAL PERFORMANCE

by

Kelly Ann Wills

Submitted to the Department of Economics

in partial fulfillment of the requirements for

the degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

August 2012

Abstract

There have been multiple studies performed examining the effect that participation in extracurricular activities has on various aspects of people's lives, including educational performance. The process of selection bias as well as the potential correlation of the error terms has not been addressed with respect to this issue. Through the use of the Heckman two step procedure and a bivariate probit model this study provides additional information on this relationship. Different types of activities are also included. When controlling for several aspects impacting whether an individual will participate, as well as how they will perform academically, a significant level of selection bias has been found in each specification used. The error terms between the participation and grade equations are found to be correlated justifying the use of the bivariate probit model. In the Heckman selection model, the treatment effects are found to vary in sign across specifications. The bivariate probit model produces significant positive treatment effects in each specification except when considering the 'Artistic Participants'.[1]

[1] I would like to thank my supervisor Marco Cozzi. His guidance, support, and encouragement helped immensely in the production of this essay.

Table of Contents

List of Tables

Introduction:

Extracurricular and community based involvement provide a wide variety of positive outcomes for the individuals who participate in them. These activities aid in personal and social development, as well as allow for connections to be made between various people in the community and school environments. The effect of participation on assorted aspects of academic performance has been examined; however whether there is a level of self selection bias present and its direction has yet to be studied. If selection bias is present, any estimate of the effects of participation on academic performance will be biased. This paper will use a variety of community based involvement as well as school extracurricular involvement to examine if selection bias is present. The potential correlation between the error terms will be addressed as well. The treatment effect that participation has on individuals will also be calculated.

There are various positive qualities that a child or young adult may develop by participating in school extracurricular and community based activities that will also promote academic achievement according to Snyder and Spritzer (1990). They will have increased exposure to, and attention from, people within their community that may not have occurred otherwise, thereby enhancing their social skills and interactions. The people in the community and school environments may take an interest in the participant's lives, including their academic performance. By participating in team sports, group lessons, or clubs, teamwork and the ability to work under the supervision of a coach or mentor have the opportunity to develop. Participation in extracurricular activities may also increase the individual's interest in school and improve their academic

performance through eligibility requirements. Success in participation endeavors could also promote a higher level of self-worth, encouraging individuals to perform better in other areas of their lives, including academics. The possibility or dream of participating at the post secondary level may also help improve educational performance. Lastly, participation in extracurricular and community based activities may also encourage peer interactions with others who are interested in academic success.

Hanks and Eckland (1976) also provide justifications as to why participation encourages a higher level of academic performance. They maintain that participation will generate and reinforce any educational goals through exposing individuals to social networks of teachers and peers which serve as a binding element between individuals and schools, as well as the school structure. Secondly, participation facilitates achieving these goals by providing individuals the opportunity to develop the necessary knowledge, self confidence, and interpersonal skills. On the other side of the spectrum, participation in these activities requires time after school and on weekends that some could argue might be better spent studying or completing the required assignments for school. These positive and negative effects are being experienced simultaneously. This being the case, the overall outcome will be interesting and provide a look into how these activities are impacting educational performance.

The studies that have been conducted examining extracurricular involvement and academic performance have found that participants experience positive outcomes in their academic career. Participants tend to have higher grades, reduced dropout rates, and a higher rate of college graduation. These studies have mostly been conducted using data

from the U.S., and neglect the fact that self selection bias may be present. I aim to contribute to the existing literature in the following ways. By using Canadian data, I hope to determine whether similar relationships can be identified that have been shown in the American and German studies. Through examining participation in a broader sense, by including community involvement as well as extracurricular involvement, I hope to determine whether the positive effects that extracurricular activity participants have been shown to experience will still be present. By breaking participation down into two types of activities I will be able to determine whether there are any differences in direction or magnitude between the activity types. Lastly, by relying on the Heckman two step procedure and a bivariate probit model I will be able to break the analysis into parts thereby identifying key factors that influence participation and academic achievement. I will also determine whether there is any selection bias or correlation of the error terms occurring, and calculate the treatment effect experienced by participants correcting for these issues.

Literature Review:

Many studies have been done to examine the effects that participation in sports and other activities has on various aspects of people's lives, one being employment and labour market outcomes. Lechner (2009) examined long run labour market outcomes and the effect that participation has on these and other personal attributes. He found that participation provides significant positive labour market effects in the long run. Participants tended to experience higher wages and earnings, and levels of health and subjective well being. It has also been demonstrated by Cabane (2011) that the probability of an individual becoming employed is correlated significantly with the act of participating in weekly sporting activities during their unemployment.

In a study performed by Duda and Nicholls (1992), the underlying aspects of motivation for performance in both sports and academics are examined in order to determine whether any similarities or differences exist. Using factor analysis based on a number of questions about school and sport involvement, and achievement, they determined that there are four factors found to be significant in the process of determining success, and that they were identical for both the sport and academic domains. The first is cooperation as a goal, defined as the belief that effort and collaboration is important in producing success. The second significant factor is the belief that the use of deceptive tactics is necessary in order to be successful. Next, ego orientation is found to be a significant motivation. This represents the belief that the ability to become superior or to defeat others is crucial to being successful in both the sport and academic domains. Lastly, they found that there was a strong link between alienation in school and alienation

in sports. This shows that "conceptually similar achievement goals and beliefs [are] fairly closely associated, and the dimensions cut clearly across the two achievement settings" (Duda and Nicholls, 1992). In addition, they examine the correlations that exist in the sport and school domains and compared them. They find no differences in correlations between the sport and academic domains except for a higher correlation between perceived ability and satisfaction in the sport domain. Knowing that the significant motivations for achievement in academic and sport endeavors are similar, it is possible to speculate that performance in the two areas may be linked. This shows that participation in sports may be developing the necessary motivations deemed significant in obtaining higher grades in schooling, or vice versa.

With this link developed between school and sport involvement, one might ask why this exists. The notion of social capital was first introduced by Coleman (1988) and represents the connections a person exhibits between their family as well as others in their community. Using moving as well as changing schools, Pribesh and Downey (1999) examine the effects on measures of social capital in addition to their academic performance. They break down social capital into measures of student-school and student-community connections. In both of these measures they use involvement in a wide variety of school and community based activities. Pribesh and Downey find that moving, as well as changing schools, decreases a student's level of social capital. Using this fact, they then examine whether there is a significant impact on a student's academic performance. As demonstrated by lower scores on math and reading tests administered by the National Center for Educational Statistics, they find that there is also a decline in educational performance. This study uses longitudinal data obtained by the National

Educational Longitudinal Study and consists of a very large sample size. One shortcoming of the data is that there are large amounts of attrition due to relocation, and these were the students found to be the most likely to move during the original data collection year. However, the findings of the study would most likely increase in magnitude if these individuals were included throughout.

After discussing the relationships between extracurricular and community involvement and academics, as well as why this may exist, it will be useful to examine many other studies that have been conducted on the subject at hand. These have mostly been conducted throughout the U.S. and one was done in Germany. These examine the impact that participation in extracurricular activities may have on the educational performance of youth. The effects on educational performance, risky behaviours, college attendance, and high school dropouts are some of the major issues that are considered with respect to extracurricular activity involvement.

Carnelißen and Pfeifer (2010) use data from the German Socio-Economic Panel to examine whether activity or competitive participation has an effect on secondary graduation. In Germany, secondary degrees are divided into four categories based on their level: no degree, low degree, intermediate degree, and high degree. They also examined the genders separately to determine whether these effects differ between male and females. They find that the probability of obtaining the highest degree level is larger for both male and female participants, with an increase of 6.1% and 5.6% respectively. There is no difference in likelihood between participants and nonparticipants for the intermediate degree level. A lower likelihood is found for the lowest degree level with

male participants being 6.6% less likely and females 11%. When examining participating competitively they find no significant effect for males, but did find that the change in likelihoods for female participants is significant. There is a lower likelihood for female participants of obtaining the lowest degree level of 5.6%, and an increase in likelihood of 6.6% of obtaining the intermediate degree level. There is no significant difference in probability of obtaining the highest level of degree between competitive participants and nonparticipants, which Carnelißen and Pfeifer attribute to the time constraint the individuals would face.

In a study done by Eccles et al. (2003), the effects of various types of extracurricular activities are examined using data obtained by the Michigan Study of Adolescent Life Transitions. The longitudinal data follows approximately 1200 individuals and begins when they are in grade ten. They break down extracurricular involvement into the following categories: pro-social, team sports, performing arts, school involvement, and academic clubs. Participation in these activities is used to determine whether there is an effect on risky behaviours, grade 12 GPA, high school and college graduation, and enjoyment of school. They find that for all activity types, students who participated were predicted to have a higher grade 12 GPA. They also report that participants were more likely to graduate secondary school as well as college. Participation in team sports increases the likelihood of being involved in risky behaviours including using drugs or alcohol, as well as skipping school. It is also found that being involved in prosocial activities, team sports, performing arts, school involvement and academic clubs led to a greater enjoyment of school in grade 10. Participants in performing arts, school involvement, and team sports are also found to have a higher level

of school enjoyment in grade 12. A potential shortcoming of the dataset used is that all observations are from a small area, namely southeastern Michigan. This may not be an accurate depiction of the desired population. It is also discussed that participants in each activity had positive outcomes but their magnitudes relative to each other is not discussed.

McNeal (1995), and Mahoney and Cairns (1997), find that participation in extracurricular activities decreases the dropout rate of high school students. Although not directly measuring academic performance, the effect of participation on dropout rates provides an additional layer to the analysis, and contributes important information towards the question being addressed. McNeal (1995) conducted a study of the effects on dropout rates using data from the National Center for Educational Statistics. When examining the activities and their effects individually he found that participation in athletics and the fine arts reduces the number of dropouts, and participating in academic clubs reduces the dropout rate as well; however this estimate was found not to be statistically significant. These results are obtained after controlling for factors such as race, sex and language spoken. He also examines the combined effects of participating in multiple activities. When doing so, only participation in athletic activities remains statistically significant in lowering the dropout rate. This study uses a large sample consisting of students from numerous public schools which would likely provide an accurate depiction of the population. Mahoney and Cairns (1997) use longitudinal data from youth mainly located in the U.S. They used cluster analysis to examine the effects of extracurricular activities and the relationship that participation has with various other factors. They find that the individuals in their sample that dropped out of school participated in fewer activities throughout the time that they were studied. Students were

also divided into three clusters: high, moderate, and low risk, representing their risk level of dropping out of school. An individual's cluster is determined based on their scores in the categories of socioeconomic status, grades retained, aggressive behaviour, academic performance, and popularity with peers. They find that students who were at a low risk of dropping out were not significantly impacted by participation in activities, but those students who were determined to be high risk were significantly impacted. The sample used for this analysis consisted of 392 individuals from various schools across the nation. With a sample size this small, the analysis may not provide an accurate depiction of the population in question.

As the previous studies have shown, there does seem to be a positive effect that many students experience in their grades and high school experiences resulting from participating in extracurricular activities. But one thing that has been neglected to be accounted for is the selection bias that could be impacting the results of the previous studies. In Eitle and Eitle (2002), the reasons as to why males may participate in basketball, football, and other sports, is examined. The category of other sports includes a variety of sports and is included generally as when they were split into the different sports they did not provide any additional insight. The downfall of this study is that female participation rates are not examined, and only black and white males are included, resulting in key demographics missing from the analysis. They find that certain characteristics seem to have an impact on participation rates of the various sport groups. A household's socioeconomic status has an impact on participation rates for basketball and other sports, but not football. An increase in the number of cultural trips and classes decreases participation rates for basketball. Increasing the number of cultural trips

decreases the participation rates for football. Black males are more likely to participate in basketball and football, and less likely to participate in other sports. As the proportion of minorities in a given area increases, participation rates in basketball and football decrease. They also find that if a black male is doing poorly in school they are more likely to turn to participating in sports. They then examine the effects of participation on academic outcomes of both standardized tests given, as well as a measure using self reported grades. Eitle and Eitle find that when examining the standardized test scores, those that participate in football or basketball experience lower scores regardless of their race. Those that participate in the other sports were found to have higher scores. With regards to the self reported grades measure they find that white males that participate in other sports experience a higher grade measure, and black males that participate in other sports experience a lower grade measure.

As shown, previous studies have found that extracurricular activity participation tends to have a positive relationship with school performance in youth. Not only does participation seem to decrease the number of students dropping out, it increases enjoyment of school as well as grade 12 GPA's. However, there have also been findings of lower scores and self reported grade measures for participants. There also seems to be various characteristics that affect who chooses to participate in the various forms of involvement. I hope to address the issues I have discussed in the studies presented and provide a look into how community and school involvement impacts grades, accounting for the selection bias that may be occurring. I will be using a large cross section of individuals from across Canada to address the small sample size and limited area of selection that has been present in past studies.

Data and Methodology:

The National Longitudinal Survey of Children and Youth (NLSCY) will be used to conduct this analysis. The NLSCY follows children from early ages and "is designed to collect information about factors influencing a child's social, emotional and behavioural development and to monitor the impact of these factors on the child's development over time" (Statistics Canada). The study began in 1994 and has since been administered through eight cycles. The sample that will be used are 16 to 17 year olds that are currently in school in order to represent the population of high school students. This requires using cycles four through eight, and consists of approximately 7,000 individuals. Collection methods utilized by the NLSCY vary depending on the variable. Certain measures are self reported, while others are reported by the primary care giver of the youth.

The activity dummy variable is self reported by the individual. The activities that I will be using are mainly community based, with one being school based. Respondents were asked how often they participated in the activity during the past twelve months. The restriction of outside of school was included for the community based activities. The activities that will be included in the general participation variable are sports or physical activities with a coach or instructor (organized sports), art, drama or music lessons, gymnastics, dance or other group lessons, community, political or religious groups (community clubs), and school clubs, teams or other school organizations (school clubs). Respondents with any level of participation were included as participants, and non-participants were those that responded with never participating. In addition, participation variables were created for different types of activities. First is 'Artistic Participants',

including art, drama or music lesson participants, and gymnastics, dance or other group lesson participants. Second is ‘Team or Club Participants’, including organized sport participants, community club participants, and school club participants.

The variable grade is also self reported by the individual. The respondent was asked about their current grade average and was given a scale. This scale varies between the cycles. Cycle four and five respondents were given a scale using terms between ‘very well’ and ‘poorly’. Cycles six through eight were given a scale of grade values between ‘90% to 100%’ and ‘Less than 50%’. The cycles were aggregated and grade values were given a subjective term value according to table 1. As I do not know how these students may describe their grade values in descriptive terms, this could potentially pose a problem in the analysis. The coded value of each grade category is also presented in table 1.

Table 1: Grade Value and Subjective Term Conversion

Coded Value	Descriptive Term	Grade Value
1	Very Well	90%-100%
2	Well	80%-89%
3	Average	70-79%
4	Poorly	60%-69% 55%-59% 50%-54% ≤ 50%

For the bivariate probit model, the grade variable is modified into a dummy variable. The categories of ‘very well’ and ‘well’ are coded as one and represent 50.02% of the sample. The other 49.98% of the sample consists of the categories ‘average’ and ‘poorly’ and are coded as a zero.

The makeup of the dataset provides insight into the sample being used. The distribution of the grades are what one would expect with 14.3% in ‘very well’, 35.6% in ‘well’, 35.8% in ‘average’, and 14.3% in ‘poorly’. There are a higher proportion of females in the ‘very well’ and ‘well’ categories, and males in the ‘average’ and ‘poorly’ categories. The natural logarithm of household income is used and varies slightly between the grade categories with individuals in the ‘well’ category having the highest average household income, followed by ‘very well’, ‘average’, and ‘poorly’. The number of part time job hours also varies between the grade categories increasing as you move from ‘very well’ towards the ‘poorly’ category. The average amount of time spent studying and working on assignments outside of school hours is 6.7 hours. The means and standard deviations can be found in table A1. The individuals who have failed a course are more likely to be in the ‘poorly’ and ‘average’ grade categories. Most of the sample, 68.9%, have parents who are always willing to help with school. A large portion of the sample lives with two parents at 73.9%, and the amount of individuals living with one parent and independently is 22.8% and 3.3% respectively. Approximately 28% of the sample have moved in the last two years, while 20% have experienced a school change in the last two years. Only a small portion of the sample experiences some sort of difficulty, 10.8%, defined as a difficulty hearing, seeing, communicating, walking, climbing stairs, bending,

learning or doing similar activities. The proportion of individuals with a difficulty increases as you move towards the lower grade categories.

The participation rate for the general participation variable is 76%, 38% for the 'Artistic Participants', and 72% for the 'Sport and Club Participants'. The proportion of individuals who participate decreases if they have experienced a move in the last two years. Regarding household structure, individuals in a two parent household have a higher rate of participation, and those living with one parent or those living independently have a lower rate of participation. The natural logarithm of household income varies slightly between participants and nonparticipants for each of the activity types. But each activity category experiences a higher average household income for participants. Dance, gymnastic, and other group lesson participants have the highest average household income. These results are presented in table A1. The proportion of individuals who participate with access to a car is much higher than those who do not have access. The average number of older siblings is 0.52, and younger siblings is 0.57. The number of part time job hours vary between participants and nonparticipants and is also presented in table A1. A large portion of individuals in the sample are living in an area with a population of greater than 500,000, with 41.9%. The proportions in the remaining categories of residence size are similar in magnitudes. The correlations between several of the continuous variables can be found in table A2. The only variables found to be highly correlated are older and younger siblings with a negative correlation, and the number of school changes and moves with a negative correlation. Household income is moderately correlated with school changes (negative), older siblings (positive), and number of moves (negative).

Heckman Two Step

A Heckman two step model will be utilized in this analysis. Heckman (1979) introduced the two step method to account for selection bias that may be present due to self selection by the individuals or by the data selection methods. Since the former may be present in the act of selecting to participate in activities, this model is ideal for examining the relationship between grades and participation in extracurricular and community activities.

In the first step, the selection equation, a probit regression is performed with the following specification:

$$Participation_i = \Phi(x_i' \cdot \beta + \varepsilon_i)$$

where participation is a dummy variable representing whether person 'i' participates in any of the activities, β is a vector of parameters for x_i, a vector of control variables measured for each individual 'i'. Lastly, ε_i represents the error term for the individual 'i'. Φ represents the cumulative density function. Next, the selection correction term named λ_i is created according to the Inverse Mills Ratio as follows:

$$\lambda_i = \frac{\Phi(z_i)}{\phi(z_i)}$$

where Φ is the density function and ϕ is the distribution function, both for a standard normal variable, and

$$z_i = \frac{-\beta_1 \cdot x_i}{\sigma^{\frac{1}{2}}}$$

In stage two, the outcome equation, an ordered probit regression will be performed on participants and non-participants separately using λ_i as an additional variable. The four

possible grade outcomes being 'very well', 'well', 'average', and 'poorly'. This will provide insight as to whether there is a bias resulting from individual selection. The regression will be of the form:

$$Grade_i = \Phi\ (w_i' \cdot \beta\ +\ \lambda_i \cdot \alpha +\ \varepsilon_i)$$

where grade is the self reported grade variable measured for each individual 'i', β is a vector of parameters for w_i, a vector of controls also measured for each individual. α is the parameter for λ_i, and ε_i is the error term for individual 'i'. Φ represents the cumulative density function.

Bivariate Probit Model

A bivariate probit model allows us to model two equations in which the error terms may be correlated (Greene, 2012). The equations will be as follows:

$$Participation_{1i}^{*} = x_{1i}' \cdot \beta_1 +\ \varepsilon_{1i}$$

$$Grade_{2i}^{*} = x_{2i}' \cdot \beta_2 +\ \varepsilon_{2i}$$

where participation and grade are as described above with the subscripts representing the equation number.[2] The terms x_1 and x_2 are the vector of controls for the participation and grade equations respectively, measured for each individual 'i'. β_1 represents the vector of parameter estimates for the participation equation, and β_2 represents the vector of parameter estimates for the grade equation. ε_1 and ε_2 are the error terms described by

$$\begin{pmatrix}\varepsilon_1 \\ \varepsilon_2\end{pmatrix}\Big| x_1, x_2\Big) \sim \mathrm{N}\left[\begin{pmatrix}0 \\ 0\end{pmatrix}\ ,\ \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}\right]$$

[2] Equations with a * refer to the underlying latent variable

where ρ represents the correlation between the error terms of the two equations. If ρ is statistically significant, the error terms are correlated and the two equations must be estimated jointly.

The following sections describe the variables that will be used for both the Heckman two step procedure and the bivariate probit model. The first section outlines the variables in the selection equation for the Heckman two step, and the participation equation for the bivariate probit. The outcome equation variables for the Heckman two step will also be used as the variables in the grade equation for the bivariate probit model.

Selection Equation – Participation:

A large number of variables will be utilized in the selection equation. These are the variables that are believed to have an impact on whether an individual participates in any of the activities. First, a gender dummy variable is included with female respondents coded as a one and males as zero. Household income is reported by the primary caregiver of the individual. The value obtained is imputed using answers that were given for other questions regarding income from various sources. The natural logarithm is applied to household income in order to account for the skewed distribution of income. Household structure is a measure of the number of caregivers of the individual. Two parents are coded as zero, one parent as one, and no parents or living independently as two. These categories include biological, adoptive, step and foster parents. Next, variables present in the data set that could have an impact on participation will be discussed. Self reported part time job hours per week is included as the number of hours an individual works, the less time they would have to participate in the activities. These hours include any paid or

unpaid work for the family business or around the house, hours at any other business, as well as coop or work placement hours. Whether the individual experiences any difficulties hearing, seeing, communicating, walking, climbing stairs, bending, learning, or doing similar activities is also included as a dummy variable with a one representing that the individual does in fact experience any of the difficulties. Any of these could present an individual with difficulty in participating. Next, a dummy variable measuring whether an individual has moved in the last two years, with a zero representing no move, and a one representing one or more moves is included. Since a large portion of the activities are community based, whether an individual has moved may impact participation rates. Additionally, the factors addressed by Pribesh and Downey (2002) regarding residential moves and social capital justify the inclusion. The size of area of residence is also included and is between one and five depending on the area where the respondent lives, increasing with the population. The justification being that an individual living in a large urban area will have a much higher variety of activities available than an individual living in a rural area. The number of older and younger siblings is then included. The number of siblings in a household may have bearing on participation in community or school activities due to the justification that a higher number of siblings may imply a lower share of the resources of the household (Carnelißen and Pfeifer, 2010). Next, whether the individual has access to a car is included as a dummy variable. Whether they are able to travel easily to these different types of activities would affect participation rates. Lastly, cycle dummy variables are included to account for the fact that these variables are measured at different periods of time.

Outcome Equation – Grade:

Several control variables are included in the outcome equation including the observables as described previously in the selection equation. These are gender, household income, and household structure. In addition, part time job hours and the dummy variables for moving, difficulties, and the different cycles will also be included. These are also measured as described previously. New to the outcome equation is a school change dummy coded as a one if an individual has changed schools for any reason in the past two years. Changing schools may have an impact on a student's grade for various reasons including the loss of social capital as discussed by Pribesh and Downey (2002). Next, whether the individual has failed a course in the past two years is included as a dummy variable. Whether a student has failed a course in the past two years may be an indicator as to how they perform in later years. A variable representing the level of parental help is then included. This variable is self reported varying from zero to five. A zero represents a student never needing any help from their parents, and increases as the level of parental help lowers from always willing to help, will help most times, will help sometimes, will help rarely, and will never help. Lastly, time spent studying is included. The variable represents the amount of time the respondent spends studying and doing assigned work outside of regular school hours. It is self reported and ranges between zero and forty hours. Although the respondents are given multiple tests and assessments, the tests given to the different cycles vary both in type and scoring. These would be preferred as a control for ability; however, as the tests in each cycle are different, it would be problematic using these.

One important aspect of the Heckman two step procedure is the exclusion restrictions (Greene, 2012). These are the variables that are present in the selection equation but not in the outcome equation. They allow for credible estimates to be generated provided that at least one is statistically significant. In specification one, and in the 'Artistic Participants' and 'Sport Participants', the exclusion restrictions are the size of the area of residence, older siblings, younger siblings, and access to a car. [3]

The size of the area of residence is included in the selection equation to account for the different opportunities an individual would have in the different locations. There would be much more options for an individual living in an area with a population greater than 500,000 than an individual living in a rural area. However, in each of these area sizes there will be schooling opportunities for each individual that will be following provincial standards. This is why it is included as an exclusion restriction. Next are older and younger siblings. These are included in the selection equation as each additional older or younger sibling implies that each one receives less of the family share (Carneliß en and Pfeifer, 2010) which may mean that there are potentially less opportunities to participate. The presence of older or younger siblings does not however reduce opportunities to attend school. Lastly, access to a car is included as participation may be affected by whether one has transportation to the activities. Many of the activities included are community based and transportation would not be provided. Access to a car would not have an effect on academic performance as transportation to and from schools is usually provided if it is needed.

[3] These are also the exclusion restrictions in specification five presented in the appendix

Specification two has three exclusion restrictions: access to a car, whether the individual has moved, and part time job hours. The justification for the inclusion of access to a car is mentioned above. Whether the individual has moved is included as per the discussion provided by Pribesh and Downey (2002). They found evidence of decreased social capital in the form of participation in community based activities as a result of moving. The number of part time job hours is the last exclusion restriction present in specification two. Since many of the activities included are community based, this would imply that they would take place in the evenings and on weekends. This would also be the times when individuals would be able to work a part time job, making participation more challenging as the number of part time job hours increases. But working a part time job would not impact the hours an individual spends at school.[4]

[4] The exclusion restrictions in specification six found in the appendix are part time job hours, whether the individual has moved, the number of younger siblings, and access to a car. The justifications for each of these restrictions have been discussed and apply to this specification as well.

Results:

A preliminary analysis was performed to examine whether a relationship could be found between participation in extracurricular and community activities and an individual's grade, and if it was a positive or negative relationship. Using an ordered probit model controlling for gender, household income, the size of area of residence, and the time spent outside of school hours doing school work, and examining the marginal effects, it was found that participants in each of the five activity categories were more likely to obtain grades in the categories of 'very well' and 'well', while they were less likely to obtain grades in the 'average' and 'poorly' categories. Although marginal effects for each of the four categories were small, it does support the findings of other studies that have been performed. These results were significant at the one percent level for all activity categories except the dance, gymnastics, or other group lessons category which was significant only at the ten percent level. These results are shown in tables 2 and 3.

Table 2: Preliminary Ordered Probit Analysis

Variable	**Coefficients**				
Organized Sports	-0.138***				
	(0.042)				
Art, Drama, or Music Lessons		-0.197***			
		(0.048)			
Gymnastics, Dance or Other Group Lessons			-0.098*		
			(0.052)		
Community Clubs				-0.142***	
				(0.049)	
School Teams or Clubs					-0.139***
					(0.042)
Gender	-0.221***	-0.198***	-0.199***	-0.202***	-0.204***
	(0.042)	(0.043)	(0.043)	(0.043)	(0.042)

Size of Area of Residence	0.018 (0.014)	0.019 (0.014)	0.020 (0.014)	0.017*** (0.014)	0.020 (0.014)
Household Income	-0.091*** (0.035)	-0.100*** (0.035)	-0.103*** (0.035)	-0.103*** (0.035)	-0.104*** (0.035)
Time Spent Studying	-0.042*** (0.004)	-0.041*** (0.004)	-0.042*** (0.004)	-0.041*** (0.004)	-0.042*** (0.004)

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

Table 3: Preliminary Marginal Effects

Activity	**Very Well**	**Well**	**Average**	**Poorly**
Organized Sports	0.029	0.027	-0.026	-0.029
Art, Drama or Music Lessons	0.043	0.036	-0.039	-0.039
Gymnastics, Dance or Other Group Lessons	0.021	0.018	-0.019	-0.020
Community Clubs	0.030	0.026	-0.028	-0.029
School Teams or Clubs	0.029	0.027	-0.026	-0.029

Heckman Two Step Procedure:

The Heckman two step procedure was then carried out with various specifications. The first specification includes all of the variables discussed above for both the selection and outcome equations. The second specification includes a subset of the variables. The third and forth specifications are the 'Artistic Participants' and 'Sport and Club Participants' respectively, and include each of the variables included in specification one. The results of the selection equation specifications can be found in table 4 and the results of the outcome equation specifications can be found in table 5.[5]

[5] In addition, a few other specifications were conducted in order to test robustness of the model and are included in the appendix in tables A3 and A4 displaying the selection and outcome equation outputs respectively

Table 4: Selection Equation Output

	Selection Equation			
Variable	**Specification 1**	**Specification 2**	**Specification 3: Artistic Participants**	**Specification 4: Sport and Club Participants**
Gender	0.156***	0.155***	0.371***	0.085
	(0.054)	(0.054)	(0.051)	(0.053)
Household Income	0.180***	0.206***	0.142***	0.182***
	(0.054)	(0.052)	(0.048)	(0.052)
Household Structure	-0.065	-0.071	-0.074	-0.048
	(0.076)	(0.075)	(0.069)	(0.074)
Part Time Job Hours	-0.002	-0.002	-0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Difficulty	-0.019	-0.034	0.087	-0.028
	(0.081)	(0.082)	(0.080)	(0.080)
Move Dummy	-0.280***	-0.297***	-0.157**	-0.256***
	(0.079)	(0.080)	(0.070)	(0.075)
Size of Area of Residence	0.040**		0.047***	0.022
	(0.018)		(0.017)	(0.018)
Older Siblings	0.094**		0.022	0.135***
	(0.044)		(0.039)	(0.042)
Younger Siblings	0.075**		0.018	0.088***
	(0.035)		(0.029)	(0.033)
Access to Car	0.359***	0.331***	0.134**	0.366***
	(0.055)	(0.054)	(0.528)	(0.054)
Cycle Five Dummy	0.118	0.110	0.025	0.174**
	(0.078)	(0.008)	(0.081)	(0.077)
Cycle Six Dummy	0.417***	0.405***	0.155*	0.389***
	(0.082)	(0.083)	(0.080)	(0.081)
Cycle Seven Dummy	0.546***	0.548***	0.278***	0.542***
	(0.093)	(0.093)	(0.087)	(0.091)
Cycle Eight Dummy	0.196**	0.202**	0.082	0.218***
	(0.085)	(0.084)	(0.080)	(0.082)
Constant	-1.731***	-1.757***	-2.293***	-1.895***
	(0.638)	(0.629)	(0.560)	(0.616)
Log Likelihood	-1352737.6	-1359731.7	-1881428.9	-1511363.4

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

Table 5: Outcome Equation Output

Outcome Equation				
Variable	**Specification 1**	**Specification 2**	**Specification 3: Artistic Participants**	**Specification 4: Sport and Club Participants**
Gender	-0.237***	-0.247***	0.016	-0.247***
	(0.052)	(0.052)	(0.159)	(0.051)
Household Income	-0.007	-0.025	-0.069	-0.001
	(0.054)	(0.052)	(0.090)	(0.054)
Household Structure	0.005	0.008	-0.229**	0.034
	(0.070)	(0.071)	(0.101)	(0.073)
Part Time Job Hours	0.003		-0.001	0.003
	(0.002)		(0.003)	(0.002)
Difficulty	0.201**	0.206**	0.364**	0.187**
	(0.081)	(0.081)	(0.113)	(0.082)
School Change Dummy	0.022	0.004	0.100	0.020
	(0.064)	(0.064)	(0.097)	(0.018)
Move Dummy	-0.129*		-0.180*	-0.144**
	(0.070)		(0.107)	(0.070)
Fail Dummy	0.843***	0.845***	0.813***	0.863***
	(0.058)	(0.058)	(0.082)	(0.061)
Parental Help	0.154***	0.153***	0.103***	0.177***
	(0.036)	(0.036)	(0.040)	(0.037)
Time Spent	-0.029***	-0.029***	-0.033***	-0.030***
	(0.004)	(0.005)	(0.006)	(0.005)
Cycle Four Dummy	-0.776***	-0.764***	-0.384***	-0.782***
	(0.086)	(0.083)	(0.116)	(0.090)
Cycle Five Dummy	-0.712***	-0.710***	-0.676***	-0.701***
	(0.076)	(0.075)	(0.112)	(0.078)
Cycle Six Dummy	0.195**	0.165**	0.212**	0.149*
	(0.078)	(0.077)	(0.105)	(0.078)
Cycle Seven Dummy	0.077	0.048	0.201*	0.091
	(0.078)	(0.077)	(0.116)	(0.078)
Lambda	0.836***	0.582**	1.084**	0.754***
	(0.306)	(0.296)	(0.554)	(0.272)
Log Likelihood	-2514819.9	-2519411.8	-1258471.1	-2363221.5

Cut 1	-1.130	-1.438	-1.178	-0.981
	(0.701)	(0.672)	(1.550)	(0.710)
Cut 2	0.188	-0.121	0.162	0.338
	(0.697)	(0.669)	(1.548)	(0.706)
Cut 3	1.516	1.204	1.445	1.685
	(0.695)	(0.666)	(1.543)	(0.704)
Cut 4	3.516	3.194	3.447	3.852
	(0.712)	(0.694)	(1.576)	(0.716)

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

When looking at the selection equation, we can see that many of the variables are having a statistically significant impact on the probability of participating. Several variables are consistently significant including household income, moving, and access to a car. Interestingly, gender is not significant in the 'Sport and Club Participants' specification implying that whether an individual is male or female is not a significant predictor. Household structure, part time job hours, and the presence of a difficulty do not present a significant impact on participation in any of the specifications. The presence of older and younger siblings, the size of the area of residence, and the different cycle dummies vary in significance between the specifications.

Examining the sign on the variables will give some insight into which factors are impacting participation, and in what direction. If the individual is a female, she is more likely to participate based on the positive coefficient. However, this value is not significant in the 'Sport and Club Participant' specification as mentioned prior. The coefficient values for gender are very similar for specifications one and two but are very different for 'Artistic Participants' and 'Sport and Club Participants'. An increase in household income increases the likelihood of participating in each specification and this result holds also for an increase in the size of the area of residence, and an increase in the

number of siblings whether they are older or younger. The values of the estimate for household income are similar across specifications, only lowering slightly in the 'Artistic Participants' specification. The coefficient estimates for the size of the area of residence, older siblings, and younger siblings are very similar for the statistically significant estimates and vary slightly in the specifications where they are not significant. As an individual experiences a decrease in the number of parents or caregivers, the individual's likelihood of participating decreases as would be expected, however this estimate is not statistically significant. This result also holds as the number of part time hours an individual works increases. The size of the estimates for part time job hours and household structure are very similar across specifications. If an individual has moved they have a lower likelihood of participation in each specification. These estimates are also similar in magnitude across specifications, only lowering slightly in the 'Artistic Participants' specification. The estimate on the difficulty dummy variable varies in direction between the specifications. It has a negative effect in the 'Artistic Participants' specifications, and a positive effect in the other specifications. If the individual has access to a car they are more likely to participate. The magnitude of the change is much lower in the 'Artistic Participants' specification, but the values are very similar in the other specifications. All of the directions of the impacts are what would be expected for each of the variables presented in the model, although the positive effect for gender is surprising.

Looking at the outcome equation output, we first need to examine the lambda estimate. In each of the specifications we find that the lambda coefficient is significant at the five percent level or higher implying that there is self selection occurring in the sample. This shows that merely examining the impact of participation on grades would

results in biased estimates. After accounting for the probability of participating, we can now examine the coefficients of the other variables. We see that whether the individual has any difficulty, whether they have failed a course, the level of parental help, whether they have moved, and the time they spend studying outside of school hours are significant in each of the specifications. The individual's gender is not a significant factor for the 'Artistic Participants'. Household income, part time job hours, and whether they have changed schools are never statistically significant. Lastly, the variables representing household structure and the various cycle dummy variables vary in significance by the different specifications.

The interpretation of the coefficients of an ordered probit model do not provide any direct insight into the magnitude of the effect a variable is having. However, by examining the sign of the estimate we can determine which direction the distribution will move with an increase in the variable. This will provide insight into how the densities in the 'very well' and 'poorly' categories will change. The effect experienced in the 'well' and 'average' categories will be ambiguous. A positive sign implies that an increase in the variable shifts the distribution to the right, decreasing some mass in the leftmost cell, 'very well' in our case, and increasing the mass in the rightmost cell, 'poorly' in our case. A negative sign will have the opposite effect, shifting the distribution to the left. This would mean a decrease in the mass in the rightmost cell, and an increase in the mass in the leftmost cell. (Greene, 2012)

The variable gender varies in sign, however in the specifications that are found to be significant it has a negative sign. This implies that if the individual is a female the

distribution shifts left. This is also true for an increase in household income. Household structure varies in sign between the specifications but is statistically significant in the 'Artistic Participants' specification and presents a negative coefficient. This implies that the distribution shifts left as the household structure variable increases, which represents a decrease in the number of parents or caregivers. This sign is the opposite of what would be predicted, however the other specifications present positive, but statistically insignificant coefficient estimates. The estimates for part time job hours also vary in sign between the specifications, although none are statistically significant as mentioned prior. The estimates for the difficulty variable are very similar across specifications and are positive implying a shift to the right towards the 'poorly' category. A change of schools presents a rightward shift of the distribution, while a move presents a leftward shift. These results are contradictory, however only the moving dummy is found to be statistically significant. The estimate for the fail dummy is positive, implying a shift to right towards the 'poorly' category. This result is also true for the parental help variable. This makes sense as an increase in the number value for the parental help variable represents a lower level of parental help. An increase in the amount of time the individual spends doing school work implies a shift to the left towards the 'very well' category as the coefficient estimate is negative. Lastly, the coefficient estimate for lambda is positive implying that as the probability of participation increases, the distribution shifts right. This would imply that the unobserved factors that make participation more likely tend to be associated with grades in the lower categories.

Since we know that selection bias is occurring in the sample, specification two was then run with and without the inverse mills ratio in order to see if certain variables

are picking up the bias that was identified prior. Table 6 compares the coefficient estimates with the estimates for the regression without the inverse mills ratio on the right. We see that all of the variables that are statistically significant in the original equation with the lambda term are still significant at the same significance level. The variables representing whether the child has a difficulty is now significant at a lower level, and the household income variable is now statistically significant. This shows that these variables may be picking up some of the effect that the probability of participating is having.

Knowing that we have a certain level of selection bias present we can calculate a treatment effect. If we used a standard ordered probit regression of the form

$$Grade_i = \Phi\ (\alpha_1 \cdot x_i\ + \alpha_2 \cdot Participation_i + \varepsilon_i\ )$$

we know that α_2 would overestimate the effect of the treatment, in our case participation. This is because we know that there are other factors affecting whether an individual chooses to be participate, or not participate. Since we have estimated the two stages of the Heckman two step model, we are able to calculate the effect that participation, our treatment, has on individuals. Using the method presented by Greene (2012) we can calculate this using the following equation:

$$E[\ Grade_i \mid Participation_i = 1,\ x_i\ ,\ w_i\ ] - E[Grade_i \mid Participation_i = 0,\ x_i\ ,\ w_i\ ]$$

and the results can be found in table 7. [6]

[6]The treatment effects for specification 5 and specification 6 can also be found in the appendix in table A5.

Table 6: Outcome Equation With and Without Inverse Mills Ratio

Variables	With Inverse Mills Ratio	Without Inverse Mills Ratio
Gender	-0.247***	-0.277***
	(0.052)	(0.049)
Household Income	-0.025	-0.077*
	(0.052)	(0.045)
Household Structure	0.008	0.043
	(0.071)	(0.068)
Difficulty	0.206**	0.217***
	(0.081)	(0.079)
School Change Dummy	0.004	0.025
	(0.064)	(0.063)
Fail Dummy	0.845***	0.857***
	(0.058)	(0.057)
Parental Help	0.153***	0.155***
	(0.036)	(0.036)
Time Spent	-0.029***	-0.029***
	(0.005)	(0.004)
Cycle Four Dummy	-0.764***	-0.708***
	(0.083)	(0.078)
Cycle Five Dummy	-0.710***	-0.684***
	(0.075)	(0.074)
Cycle Six Dummy	0.165**	0.125*
	(0.077)	(0.074)
Cycle Seven Dummy	0.048	-0.015
	(0.077)	(0.070)
Lambda	0.582**	
	(0.296)	
Log Likelihood	-1359731.7	-2521253.3

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

Table 7: Treatment Effect – Heckman Two Step

	Effect
Specification 1	0.120
Specification 2	-0.322
Specification 3: Artistic Participants	-0.433
Specification 4: Sport and Club Participants	0.226

Examining the results, we find that participation in specification two and in 'Artistic Participants' tends to have a negative effect on grades, while participation in specification one and 'Sport and Club Participants' tend to have a positive effect on grades. These results contradict each other; therefore don't really provide much insight especially between specification one and two as these are based on the same participants and nonparticipants. However, by examining the difference between 'Artistic' and 'Sport and Club' participants we see opposing effects. This result provides information regarding which type of activity participants experience positive effects, controlling for selection bias. Since it was found that boys who are doing poorly in school are more likely to turn to sports (Eitle and Eitle, 2002), and also that participation in sports and other activities increases the level of social capital thereby increasing academic achievement (Pribesh and Downey, 1999), the positive treatment effect found for 'Sport and Club Participants' may be supported. In addition, the negative effect for 'Artistic Participants' may be due to the fact that people who are drawn to participate in activities such as dance, art, or gymnastics are already doing well in school. In doing so, they have less time to spend on assigned work resulting in lower grade levels.

Bivariate Probit Model:

The specifications used in the Heckman two step procedure were also used for the bivariate probit model. The estimates obtained can be found in table 8.[7] The first step is to examine the value of ρ for each of the specifications, and to determine whether the value is statistically significant. In each of the specifications, ρ is found to be statistically significant at the 1% level. This implies that the error terms in each of the equations are correlated and in order to estimate each equation effectively a bivariate probit model must be used.

Examining the participation equation, the variables that are found to be consistently statistically significant are gender, household income, whether the individual has moved, and whether the individual has access to a car. These all have the expected effect on participation, with gender, household income, and access to a car having a positive effect, and whether they have moved having a negative effect. The variables representing the size of the area of residence, the number of older and younger siblings, and the cycle dummy variables vary in significance throughout the specifications. Each of these are found to have a positive effect on participation. Household structure, the presence of a difficulty, and the number of part time job hours are never significant. Comparing the estimates to those obtained in the Heckman two step procedure, we find that the results are very similar in magnitude, direction, and significance.

[7] As a method for testing the robustness of the model, various other specifications were estimated and the results can be found in the appendix in table A6

Table 8: Bivariate Probit Output

	Participation Equation			
Variable	Specification 1	Specification 2	Specification 3: Artistic Participants	Specification 4: Sport and Club Participants
Gender	0.190***	0.188***	0.382***	0.084
	(0.058)	(0.056)	(0.052)	(0.054)
Household Income	0.173***	0.197***	0.139***	0.163***
	(0.057)	(0.055)	(0.049)	(0.054)
Household Structure	-0.079	-0.087	-0.073	-0.073
	(0.079)	(0.079)	(0.071)	(0.076)
Part Time Job	-0.003	-0.003	-0.002	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)
Difficulty	-0.012	-0.027	0.094	-0.003
	(0.086)	(0.087)	(0.083)	(0.082)
Move Dummy	-0.248***	-0.267***	-0.147**	-0.247***
	(0.085)	(0.085)	(0.073)	(0.080)
Size of Area of Residence	0.035*		0.042**	0.019
	(0.019)		(0.017)	(0.018)
Older Siblings	0.092**		0.007	0.135***
	(0.045)		(0.040)	(0.043)
Younger Siblings	0.074**		0.018	0.090**
	(0.037)		(0.030)	(0.035)
Access to a Car	0.332***	0.306***	0.108**	0.337***
	(0.058)	(0.057)	(0.054)	(0.056)
Cycle Five Dummy	0.151*	0.147*	0.081	0.184**
	(0.084)	(0.083)	(0.083)	(0.081)
Cycle Six Dummy	0.401***	0.392***	0.158**	0.368***
	(0.084)	(0.084)	(0.080)	(0.081)
Cycle Seven Dummy	0.549***	0.550***	0.297***	0.528***
	(0.097)	(0.097)	(0.089)	(0.094)
Cycle Eight Dummy	0.196**	0.201**	0.099	0.197**
	(0.087)	(0.086)	(0.080)	(0.083)
Constant	-1.610**	-1.632**	-2.228***	-1.600**
	(0.673)	(0.662)	(0.564)	(0.638)

	Grade Equation			
	Specification 1	Specification 2	Specification 3: Artistic Participants	Specification 4: Sport and Club Participants
Gender	0.229***	0.228***	0.230***	0.230***
	(0.053)	(0.054)	(0.053)	(0.053)
Household Income	0.079	0.078	0.076	0.080
	(0.052)	(0.052)	(0.052)	(0.052)
Household Structure	0.005	0.005	0.005	0.006
	(0.076)	(0.075)	(0.076)	(0.076)
Part Time Job Hours	-0.001		-0.002	-0.001
	(0.002)		(0.002)	(0.002)
Difficulty	-0.169**	-0.170**	-0.162**	-0.169**
	(0.083)	(0.083)	(0.082)	(0.083)
School Change Dummy	0.095	0.094	0.085	0.092
	(0.075)	(0.074)	(0.075)	(0.075)
Move Dummy	0.006		0.010	0.006
	(0.075)		(0.075)	(0.075)
Failed Course Dummy	-0.855***	-0.855***	-0.868***	-0.855***
	(0.061)	(0.061)	(0.060)	(0.061)
Parental Help	-0.137***	-0.138***	-0.134***	-0.138***
	(0.035)	(0.036)	(0.035)	(0.035)
Time Spent	0.031***	0.030***	0.030***	0.030***
	(0.005)	(0.005)	(0.005)	(0.005)
Cycle Four Dummy	0.660***	0.666***	0.659***	0.660***
	(0.085)	(0.084)	(0.085)	(0.085)
Cycle Five Dummy	0.591***	0.592***	0.594***	0.592***
	(0.082)	(0.081)	(0.082)	(0.082)
Cycle Six Dummy	-0.104	-0.102	-0.105	-0.103
	(0.080)	(0.080)	(0.080)	(0.080)
Cycle Seven Dummy	0.063	0.0638	0.067	0.064
	(0.084)	(0.084)	(0.084)	(0.084)
Constant	-0.980	-0.980	-0.940	-0.987
	(0.648)	(0.647)	(0.646)	(0.648)
Log Likelihood	-2870540.3	-2877513.4	-3406808.6	-3018964
ρ	0.119***	0.121***	0.160***	0.116***
	(0.039)	(0.038)	(0.0331)	(0.037)

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

Next, we can look at the estimates obtained for the grade equation. Gender, the presence of a difficulty, whether the individual has failed a course, the level of parental help, the time they spend doing school work outside of school hours, and the cycle four and five dummy variables are significant in each of the specifications and have the predicted effect. Gender and the time spent studying each have a positive effect on the grades obtained. The presence of a difficulty, whether the individual has failed a course, and the parental help variable each have a negative effect on the grade obtained. Household income, household structure, part time job hours, whether the individual has moved or failed, and the cycle seven and eight dummy variables are never statistically significant.

Next, we are able to calculate the treatment effect using the predicted probabilities and the method presented in Greene (2012):

$$Prob(Participation = 1, Grade = 1|x_1, x_2) -$$

$$Prob(Participation = 0, Grade = 1|x_1, x_2)$$

$$Prob(Participation = 1, Grade = 1|x_1, x_2) = \Phi_2\,(x_1' \cdot \beta_1, x_2' \cdot \beta_2, \rho)$$

$$Prob(Participation = 0, Grade = 1|x_1, x_2) = \Phi_2\,(-x_1' \cdot \beta_1, x_2' \cdot \beta_2,\ -\rho)$$

where Φ_2 represents the bivariate cumulative density function. Using the above formula the treatment effects for each of the specifications are calculated and the results are in table 9.[8]

[8] The appendix also has the treatment effects calculated for the additional specifications in table A7.

Table 9: Treatment Effect – Bivariate Probit

	Treatment Effect
Specification 1	0.33738
Specification 2	0.33759
Specification 3: Artistic Participants	-0.02204
Specification 4: Sport and Club Participants	0.29573

The treatment effects found for specifications one and two, as well as for the 'Sport and Club Participants' are very similar in magnitude and direction with each experiencing a significant positive treatment effect.[9] These results are similar to those of the Heckman two step procedure; however the direction for specification two changes. There is found to be a very small negative treatment effect for the 'Artistic Participants' which supports the earlier findings of the Heckman two step procedure.

[9] These results are similar for the additional specifications as shown in table A7.

Conclusion:

Involvement in school and community based activities are generally seen as a positive contribution to the development and academic performance of children and youth. By aiding in the development of social skills, teamwork abilities, and even physical health of the individual, participation assists in developing many skills and abilities. However, since there are only so many hours in a day, there is the flip side in that the participation in these activities requires the contribution of time that could be used for studying and completing assignments for school. With the opposing nature of these effects, examining the overall impact of participation in school and community based activities provides insight into how participation in these activities affects academic performance.

Many studies have been conducted examining the effects of participating in extracurricular activities on academic performance and high school dropout rates. Generally, participation improves educational outcomes, increases a student's enjoyment of school, and reduces dropout rates. However, there is a certain amount of selection bias that was addressed by Eitle and Eitle (2002) that has not been addressed in the studies presented by Corneliβen and Pfeifer (2010), Eccles et al (2003), McNeal (1995), Mahoney and Cairns (1997), and Pribesh and Downey (1999). Eitle and Eitle (2002) present findings that certain types of individuals are more likely to turn to different types of athletic involvement, particularly black males who are not doing well in school. Due to this fact, there is a level of self selection bias that must be considered when examining the

relationship between extracurricular and community involvement and the level of academic achievement.

Through the use of the Heckman two step procedure, I was able to examine the factors that are impacting participation as well as an individual's grades, and determine whether there is the presence of self selection bias. Several factors have been found to significantly impact participation including gender, household income, whether the individual has moved, and whether they have access to a car. In addition, the number of older siblings, number of younger siblings, and the size of the area of residence were found to be significant in different specifications. Using the outcome equation, I found that gender, the presence of a difficulty, whether the individual has failed a course, the availability of parental help, and the time spent doing school work outside of school hours were found to have a significant impact on the grades obtained. Additionally, the variable lambda was found to be statistically significant at least at the five percent level in each specification. This shows that selection bias is present, and the positive sign indicates a rightward shift in the distribution. This implies that unobserved factors that tend to increase the likelihood of participation are associated with lower grade levels. I was also able to calculate the treatment effect for the grade levels obtained between participants and non-participants. These were found to contradict as Specification 2 and the 'Artistic Participant' model were found to have a negative treatment effect, and Specification 1 and the 'Sport and Club Participant' model were found to have a positive treatment effect.

Breaking down participation into various types, as well as having a general participation variable has added additional insight. I have shown that several factors in

each the selection and outcome equations are very similar, however certain differences do exist. For instance, the insignificance of gender in the selection equation for 'Sport and Club Participants' model and in the outcome equation for 'Artistic Participants' model. With the treatment effects of the two types of activity participants being of opposing signs, the implication exists that the two activity types may have differing impacts on the individuals who participate.

Utilizing a bivariate probit model allowed us to determine that the error terms in the participation and grade equations are correlated, allowing the joint estimation. The resulting estimates illustrated again that gender, household income, whether the individual has moved, and access to a car are significant predictors of participation. Gender, the presence of a difficulty, whether the individual has failed a course, the level of parental help, and the time the individual spends working on school assignments outside of school are found to be significant with regards to the grades obtained. These results are very similar to that of the Heckman two step procedure. The treatment effect was found to be positive for each specification with similar magnitudes except the 'Artistic Participant' specification which was found to be negative.

By determining that participating in community activities may improve educational outcomes, there are many political objectives that may be put into place to take advantage of this positive impact. If it is possible to increase the ease of participating in these activities, the government could improve the performance of youth in school. Even if the overall effect is small, increasing a student's likelihood of a better academic outcome would introduce huge potential benefits to society in the form of a higher level

of human capital. In addition, by knowing whether certain types of activities have a positive or negative treatment effect, we are able to identify what types of activities we should be encouraging children and youth to be participating in.

This paper has addressed the factors that influence how a student performs in school, how individual's decide to participate in community and extracurricular activities, and the bias that exists in this participation decision. Calculating the treatment effect has allowed us to observe the overall impact of participation. Examining this issue has provided additional information that has not yet been discussed, and provides a look into how certain policy objectives involving school and community based activities might impact educational performance.

Bibliography:

Cabane, Charlotte. "Unemployment Duration and sport participation: evidence from Germany." *Documents de Travail du Centre d'Economie de la Sorbonne* (2011): Working Paper .011.49.

Coleman, James S. "Social Capital in the Creation of Human Capital." *The American Journal of Sociology*. 94.Supplement: Organizations and Institutions (1988): S95-S120.

Carnelißen, Thomas, and Pfeifer, Christian. "The Impact of Participation in Sports on Educational Attainment: New Evidence from Germany." *Economics of Education Review.* 29.1(2010): 94-103

Duda, Joan L. and Nicholls, John G. "Dimensions of Achievement Motivation in Schoolwork and Sport." *Journal of Educational Psychology.* 84.3 (1992): 290-299.

Eccles, Jacquelynne S. et al. "Extracurricular Activities and Adolescent Development." *Journal of Social Issues*. 59.2 (2003): 865-889.

Eitle, Tamela and Eitle, David J. "Race, Cultural Capital, and the Educational Effects of Participation in Sports." *Sociology of Education.* 75.2 (2002): 123-146.

Greene, William H. *Econometric Analysis, 7th ed.* Upper Saddle River (NJ): Pearson Education Inc. 2012.

Hanks, Michael P., and Eckland, Bruce K. "Athletics and Social Participation in the Educational Attainment Process." *Sociology of Education.* 49.27(1976): 271-294.

Heckman, James J. "Sample Selection Bias as a Specification Error." *Econometrica.* 47.1 (1979): 153-161.

Lechnew, Michael. “Long-run labour market and health effects of individual sports activities.” *Journal of Health Economics.* 28.4(2009): 839-354.

Mahoney, Joesph L. and Cairns, Robert B. “Do Extracurricular Activities Protect Against Early School Dropouts?” *Developmental Psychology*, 33.2 (1997): 241-253.

McNeal, Ralph B. Jr. “Extracurricular Activities and High School Dropouts”, *Sociology of Education,* 68.1 (1995): 62-80.

“National Longitudinal Survey of Children and Youth.” http://www23.statcan.gc.ca:81/imdb/ p2SV.pl?Function=getSurvey&SDDS=4450&lang=en&db=imdb&adm =8&dis=2. Statistics Canada: Description. 10 July. 2012.

Pribesh, Shana and Downey, Douglas B. “Why Are Residential and School Moves Associated with Poor School Performance?” *Demography.* 36.4 (1999): 521-534.

Snyder, Eldon E., and Sprietzer, Elmer. “High School Athletic Participation as Related to College Attendance Among Black, Hispanic, and White Males: A Research Note.” *Youth & Society.* 21.390(1990): 390-398.

Appendix:

Table A1: Summary of Variables

	Mean	Standard Deviation
Number of School Changes	0.242	0.582
Number of Moves	0.077	0.442
Household Income	11.164	0.643
Sport – Participants	11.269	0.625
Sport – Nonparticipants	11.048	0.679
Community Club – Participants	11.236	0.608
Community Club – Nonparticipants	11.124	0.678
School Club – Participants	11.231	0.651
School Club – Nonparticipants	11.116	0.658
Art, Drama, Music – Participants	11.234	0.655
Art Drama, Music – Nonparticipants	11.126	0.662
Dance, Gymnastics – Participants	11.269	0.599
Dance, Gymnastics – Nonparticipants	11.124	0.674
Very Well	11.207	0.680
Well	11.215	0.645
Average	11.149	0.644
Poorly	11.089	0.682
Part Time Job Hours	5.696	10.109
Sport – Participants	7.789	11.432
Sport – Nonparticipants	7.471	12.395
Community Club – Participants	5.910	10.128
Community Club – Nonparticipants	6.344	11.552
School Club – Participants	6.484	10.118
School Club – Nonparticipants	5.109	10.118
Art, Drama, Music – Participants	7.414	10.764
Art Drama, Music – Nonparticipants	7.709	12.328
Dance, Gymnastics – Participants	7.233	10.709
Dance, Gymnastics – Nonparticipants	7.740	12.233

Very Well	4.646	9.246
Well	5.699	9.763
Average	5.987	10.500
Poorly	6.005	10.681
Gender	0.492	0.215
Household Structure	0.294	0.032
Difficulty	0.101	0.026
Number of School Changes	0.242	0.582
Number of Moves	0.077	0.442
Size of Area of Residence	2.606	0.551
Older Siblings	0.522	0.777
Younger Siblings	0.570	0.794
Access to a Car	0.530	0.432
Fail	0.311	0.045
Parental Help	1.342	0.985
Time Spent Studying	6.698	6.149

Table A2: Correlations

	Household Income	School Changes	Part Time Hours	Older Siblings	Younger Siblings	Number of Moves	Time Spent
Household Income	1.000						
School Changes	**-0.108**	1.000					
Part Time Hours	0.006	0.031	1.000				
Older Siblings	**0.142**	-0.033	-0.027	1.000			
Younger Siblings	-0.002	-0.012	0.073	**-0.264**	1.000		
Number of Moves	**-0.142**	**0.341**	0.011	-0.060	-0.029	1.000	
Time Spent	0.090	-0.052	-0.013	0.064	-0.013	0.062	1.000

Table A3: Selection Equation Output – Specification 5 and 6

Variable	Specification 5	Specification 6
Gender	0.114*	0.143***
	(0.063)	(0.053)
Household Income	0.168*	0.217***
	(0.088)	(0.052)
Past Household Income	0.066	
	(0.086)	
Household Structure	-0.015	-0.062
	(0.094)	(0.074)
Part Time Job Hours	-0.002	-0.001
	(0.003)	(0.002)
Difficulty	-0.095	
	(0.096)	
Move Dummy	-0.248***	-0.284***
	(0.092)	(0.078)
Size of Area of Residence	0.028	
	(0.021)	
Older Siblings	0.120**	
	(0.054)	
Younger Siblings	0.053	0.039**
	(0.039)	(0.018)
Access to Car	0.363***	0.352***
	(0.065)	(0.055)
Cycle Five Dummy	-0.045	
	(0.089)	
Cycle Six Dummy	0.270***	
	(0.089)	
Cycle Seven Dummy	0.363***	
	(0.102)	
Constant	-2.137**	-1.858***
	(0.841)	(0.617)
Log Likelihood	-960888.51	-1378977.3

Table A4: Output Equation Output – Specification 5 and 6

Variable	Specification 5	Specification 6
Gender	-0.258***	-0.235***
	(0.058)	(0.050)
Household Income	0.044	0.017
	(0.074)	(0.047)
Past Household Income	-0.046	
	(0.074)	
Household Structure	0.003	0.071
	(0.076)	(0.069)
Part Time Job Hours	0.002	
	(0.002)	
Difficulty	0.199**	0.2235***
	(0.100)	(0.081)
Change School Dummy	0.012	0.054
	(0.073)	(0.067)
Move Dummy	-0.200***	
	(0.078)	
Fail Dummy	0.854***	
	(0.063)	
Parental Help	0.145***	
	(0.043)	
Time Spent	-0.030***	-0.040***
	(0.005)	(0.005)
Cycle Five Dummy	-0.730***	
	(0.079)	
Cycle Six Dummy	0.212**	
	(0.085)	
Cycle 7 Dummy	0.102	
	(0.087)	
Lambda	1.137***	1.010***
	(0.3837)	(0.258)
Log Likelihood	-1977939.2	-735410.04

Table A5: Treatment Effect – Heckman Two Step – Specification 5 and 6

	Treatment Effect
Specification 5	-0.063
Specification 6	-0.268

Table A6: Bivariate Probit Output - Specification 5, 6, and 7

	Participation Equation		
Variable	**Specification 5**	**Specification 6**	**Specification 7**
Gender	0.145**	0.154***	0.170***
	(0.065)	(0.054)	(0.055)
Household Income	0.171*	0.246***	0.247***
	(0.095)	(0.051)	(0.048)
Past Household Income	0.070		
	(0.091)		
Household Structure	-0.037	-0.032	
	(0.098)	(0.075)	
Part Time Job	-0.003	0.001	
	(0.003)	(0.002)	
Difficulty	-0.106		-0.036
	(0.102)		(0.086)
Move Dummy	-0.227**	-0.271***	-0.276***
	(0.098)	(0.079)	(0.082)
Size of Area of Residence	0.0224		0.011
	(0.022)		(0.018)
Older Siblings	0.115**		0.057
	(0.055)		(0.042)
Younger Siblings	0.054	0.060*	
	(0.041)	(0.033)	
Access to a Car	0.328***	0.326***	
	(0.068)	(0.055)	
Cycle Five Dummy	-0.009		0.162*
	(0.094)		(0.084)
Cycle Six Dummy	0.262***		0.4574***
	(0.090)		(0.084)
Cycle Seven Dummy	0.371***		0.564***
	(0.105)		(0.097)
Cycle Eight Dummy			0.256***
			(0.086)
Constant	-2.157***	-2.088***	-2.239***
	(0.881)	(0.619)	(0.528)

	Grade Equation		
	Specification 5	**Specification 6**	**Specification 7**
Gender	0.261***	0.233***	0.268***
	(0.060)	(0.050)	(0.052)
Household Income	-0.052	0.071	0.028
	(0.081)	(0.044)	(0.046)
Past Household Income	0.148*		
	(0.084)		
Household Structure	0.015	-0.101	-0.061
	(0.088)	(0.065)	(0.071)
Part Time Job Hours	-0.001		
	(0.003)		
Difficulty	-0.112	-0.237***	
	(0.098)	(0.075)	
Change School Dummy	0.148*	-0.032	
	(0.087)	(0.070)	
Move Dummy	0.097		-0.005
	(0.085)		(0.073)
Failed Course Dummy	-0.796***		-0.865***
	(0.071)		(0.061)
Parental Help	-0.119***		-0.138***
	(0.014)		(0.034)
Time Spent	0.031***	0.041***	
	(0.006)	(0.005)	
Cycle Five Dummy	0.639***		
	(0.086)		
Cycle Six Dummy	-0.084		
	(0.083)		
Cycle Seven Dummy	0.073		
	(0.085)		
Constant	-1.246	-0.998*	0.081
	(0.821)	(0.527)	(0.554)
Log Likelihood	-2132586.9	-3238461.7	-3027110.6
Rho	0.156***	0.105***	0.127***
	(0.045)	(0.035)	(0.038)

Table A7: Treatment Effect – Bivariate Probit – Specification 5, 6, and 7

	Treatment Effect
Specification 5	0.344
Specification 6	0.334
Specification 7	0.347