

# The Impact of Student Background and Academic Performance on Future Income

---

*Prepared and Edited for Submission by Raphael Small*

*rsmall@haverford.edu*

*BA Economics, Expected May 2015*

*Original Collaborators: Raphael Small, Siran Tang, Yanxi Li, Atreya Dey*

*Haverford College*

*Advisor: Professor Richard Ball*

## **Abstract**

In this paper we examine the effects of student academic performance and socioeconomic background upon future income. To do so, data was isolated and prepared from the US Department of Education's National Longitudinal Study of the High School Class of 1972 before being analyzed using STATA 12. We chose to study five variables: Participant SAT Score, Parental Education Level, Parental Pretax Income, Participant Socioeconomic Status, and Total Income. Using regression analysis, we compared various linear models, both with and without interaction effects. Although we hypothesized that affluence and privilege would positively correlate with academic performance and income, our data analysis provides substantial evidence to reject such a notion. Overall, the results indicate that our variables had a trivial financial impact on income. In our data, we found that children of all backgrounds performed nearly the same on average when controlling for educational performance, contradicting the wealth of existing literature. "

## **I: Introduction, Statement of Research Question, and Review of Previous Literature**

Given that income inequality is an increasingly evident and well-publicized trend, the contemporary focus on the importance of education is understandable. Correspondingly, the value placed upon higher education, specifically with respect to pure cost, has increased dramatically in recent decades. Just as higher education has risen in relative price, so too has it become viewed as an equalizer in order to minimize the inherently unequal paths children are placed upon by pure luck at birth. Particularly in the United States, where employment and wages feel downward pressure due to globalization, the increased value of a college education is evident, both in terms of absolute cost and cultural norms. Within the preceding decades, the cost of attending a four year college has risen well above the rate of inflation, potentially indicating the continually rising cultural and economic value of higher education.

At present, the importance of going to college in the United States is constantly cited as a solution to overcome rising income inequality and social stratification. Numerous contemporary studies, however, have illustrated that graduating from college is far from the grand equalizer it is made out to be; indeed, the educational background and wealth of parents appear to correlate strongly with the financial success of their children. Even if the children of less successful parents graduate from college, and do so with high marks, they are still less likely to be as successful as those children lucky enough to have more affluent parents.

With this in mind, we therefore hypothesize that there will be a very strong correlation between privilege and income. With more resources to draw on for schooling, the more affluent will therefore perform better both academically and financially. Even controlling for academic performance and student background, we still expect the affluent to earn more income on average.

To test our hypothesis, we cleaned and analyzed our dataset using STATA 12, a standardized program used for data and statistical analysis. After running numerous regressions upon our data and examining the outcomes, we observed several contradictory patterns. Many such trends contradicted our initial hypotheses, and overall we failed to find a single coherent model within the data, thus motivating us to reject our hypothesis. In choosing our hypothesis, we relied upon our own intuition and a wealth of existing literature examining the positive relationship between socioeconomic indicators, student performance, and future income.

For instance, in their comprehensive study of the Harlem Children's Zone, a nonprofit educational organization dedicated to alleviating poverty within Harlem, Dobbie and Fryer (2011) cite numerous studies that found that increased school funding ("inputs") failed to correct the difference in performance between the children of more affluent parents, who consistently perform better on standardized tests, and those of parents of lesser means. In essence, the usual educational inputs do not account for educational success, and instead, such success is a result of a variety of parental variables. Dobbie and Fryer additionally endorse the potentially controversial notion that "schools alone cannot treat the problem of chronic underachievement" of children belonging less advantaged socioeconomic classes. Dobbie and Fryer also note that even the Promise Academy, a renowned and supposedly successful charter school located within the Harlem Children's Zone, failed to deliver statistically significant gains on the test scores of the majority of its low-income students.

In a related study, Das et al (2013) examined schools in Zambia and concluded that increasing funding to schools through traditional input channels did not lead to improved test performance. The fact that this finding occurred in the heart of Africa, obviously an extremely

different environment from the United States in numerous ways, points toward the universality of the lack of a strong connection between educational inputs, i.e. funding, and results.

If school resources do not dictate success, then what factors do? Hamnett et al (2007) are blunt: even when controlling for educational backgrounds, children from better off families are almost always more successful in terms of their educational performance. To this end, Hamnett et al show evidence that even when schools are permitted to select their students based on academic merit alone, children of higher socioeconomic status still achieve superior grades on average, and that overall, such privileged children tend to perform better even within the context of different schools. In short, children of more privileged families perform better than their peers on average, even upon controlling for academic setting.

At the same time, Dobbie and Fryer argue theoretically that even if a school district sees the income of its inhabitants rise, and if the quality of the schooling does not change, then education results as measured by standardized tests should not change in a statistically significant manner. In essence, the argument is that there is no link between the wealth of a school district and the academic results of its children, i.e. that only the schooling itself matters. This argument, however, neglects the fact that because public schools are generally financed by property taxes, more resources generally become available to school districts as the income of their respective residents increases. This, if anything, underlies the strong connection between socioeconomic status/income and educational achievement: wealthier districts are therefore able to input correspondingly more resources into education. Indeed, the flaw in this thought process actually emphasizes, if anything, the link between educational achievement and socioeconomic status.

Loken et al (2012) found that in the excellently controlled case of the oil-rich Rogaland region of Norway parental income trumped all else as indicators of their children's future successes. This conclusion is further supported by the Carneiro (2007) Portugal study, which additionally emphasizes the lack of correlation between school funding and educational success. Carneiro claims that based on his findings, the evidence supports a shift away from what he terms "traditional input based policies" towards one of educational success strongly based upon familial inputs, i.e. that those of greater means have children that tend to be more educationally successful and therefore more financially successful. Lastly, as one would logically suspect, Carneiro points to compelling evidence that significantly lower educational performance strongly correlates with lower future incomes, understanding that children of high means are generally the top performers and thus garner for themselves most of the positions within higher income brackets.

Moving to a more localized scope of study, the general question we seek to address is the nature of the connection between income and academic performance in the United States. Among college graduates, thereby attempting to control for the level of education, are test scores effective indicators of success, defined by income, later in life? What is the impact of wealth, income, and socioeconomic standing of one's parents on scores and income? Essentially, the root question we wish to answer is that of the interplay of income ("success" in life) and test scores ("success" in school): how well, and to what degree, do socioeconomic factors, when controlling for schooling, correlate with future income?

## **II: Overview of Data**

The data used within this study is entirely contained within the National Longitudinal Study of the High School Class of 1972 (US Department of Education, Center for Education

Statistics 1992, NLSC-72). The participants of the study were seniors in high school when the study began in the spring of 1972. The participants were then followed up in the years 1973, 1974, 1976, 1979, and 1986.

In order to simplify the data set, we narrowed down our original sample by restricting the data to contain only students that had graduated from high school in '72 and college in '76. Students were also removed from the dataset if they had not listed date of birth and/or gender in their questionnaires. All errors (such as non-responses, skips, or invalid answers) were then purged. The remaining participants have all graduated from high school and college, and on the whole, the data is virtually complete for each participant. Following data cleaning and preparation, 900 participants of the original survey remained.

The variables we selected for study and various analyses are as follows:

- 1) **SAT Score:** Standardized Achievement Test scores, with combined verbal and math sub-scores, each with a subsection maximum score of 800. Because of its highly standardized nature, SAT scores were used as a measure of potential academic achievement.
- 2) **Parent's Education Level:** Educational degrees or the lack thereof of the participants' fathers and mothers. This is ordered categorical data regarding the parents' degree of education as follows: low (no college experience); medium (at least some college experience); and high (at least some graduate school experience).
- 3) **Parents' 1972 Pretax Income:** In NLSC-72, the income of participants was listed within brackets. The pretax income variable for this study was obtained by estimating parental income by using the midpoint of each bracket.
- 4) **Socioeconomic Status:** This is an ordered categorical variable with values of low, medium, and high with regard to the participant's socioeconomic status. The socioeconomic status

variable is a combined total measure of a family's relative income, parental occupational prestige, and parental education level. A score of a 1 indicates a “low” socioeconomic status, lower class. A score of a 2 indicates “medium” socioeconomic status, i.e. middle class. A score of a 3 indicates high socioeconomic status, upper class.

- 5) **Total Income for 1976, '77, '78, and '79:** This variable was obtained by summing the participants' incomes in each of the listed years without accounting for inflation. The total four-year income was chosen to minimize volatility between each annual income; by adding the four together the aim was to achieve lower variation with increased “sampling.”

### **III: Regression Analysis**

As previously stated, the purpose of this study is to investigate the effects of academic performance and participant background upon income, a basic linear relationship with the interaction effect modeled by the following:

$$I = \beta_0 + \beta_1\rho + \beta_2\omega + \beta_3\rho\omega + \varepsilon$$

Where I is the total income as measured by the sum of incomes in years '76-'79 inclusive; each  $\beta$  is a constant;  $\rho$  is the chosen measure of a participant's academic performance;  $\omega$  is the chosen measure of a participant's chosen background variable; and  $\varepsilon$  the random error associated with the model, of which the usual assumptions of  $E(\varepsilon)=0$  and  $VAR(\varepsilon)=\sigma_\varepsilon^2$  apply for all  $\varepsilon$ . Clearly, with the inclusion of the fourth term we intend to regress using the “interaction effect” in order to capture the interplay of performance and background in turn when examining income. The relationships modeled by interaction effects are the primary forms of analysis for this study, and such results are of the most interest. The partial derivatives of the basic model are thus given by:

$$(1) \quad \frac{\partial I}{\partial \rho} = \beta_1 + \beta_3\omega$$

$$(2) \quad \frac{\partial I}{\partial \omega} = \beta_2 + \beta_3 \rho$$

Here, Equation 1 and 2 model the impact of student performance and background respectively upon income as single variable linear functions. With Equation 1, the impact of student performance on later income is being modeled while controlling for student background; Equation 2 is the opposite, modeling the effects of student background on income, controlling for student performance. We choose to use SAT scores as the primary measure of student performance ( $\rho$ ) because it is both standardized and continuous, making the results of regression analysis much more transparent and concise.

To begin, the most obvious regression to run is a linear fit of SAT scores to participant income (See Regression 1, Table 1). At the 0.05 level, the linear relationship is statistically significant, yielding the following relationship:

$$\frac{\partial I}{\partial SAT} = -13.20505$$

This means that for every 100 point gain on the SAT, the model predicts that the participant will lose \$1,320.505 from his/her four-year income. Considering the median value of total income is \$42,085 this 3.1% income loss over the four-year income is nearly insubstantial with the additional reflection that a 100 point gain on the SAT is a huge improvement. In fact, a 100 point gain on the SAT is actually equivalent to 1.88 standard deviations as calculated from the dataset.

In addition, the dual impact of SAT scores and parental income level upon participants' later income is another relationship we wish to examine (See Regression 2, Table 1). In doing so, we not only see that the parental income term is statistically insignificant, but also that the statistically significant SAT term is almost identical to that of the simple linear relationship

between SAT and income. Here, there is also a negative relationship between SAT scores and income according to the following derivative:

$$\frac{\partial I}{\partial SAT} = -13.64742$$

Similarly, for every 100 point gain on the SAT, this model predicts that the participant will lose \$1,364.742 from his/her four-year income. This amounts to a 2.9% income loss over the same period, a minor change considering the decrease is spread over four years.

For the interaction effect between SAT scores and parental income, we use the (n-1) dummy variable 0/1 model (See Regression 3, Table 1). Within this particular regression model, SAT scores alone are not significant at the 0.05 level in accordance with the p-value. However, at this level both parental income ( $\omega$ ) and the interaction term are statistically significant, allowing us to exploit the relationship denoted by Equation 2, in this case controlling for SAT scores. In doing so, we take the derivative with respect to parental income  $\omega$ , eliminating the statistically insignificant  $\rho$  term:

$$\frac{\partial I}{\partial \omega} = 2.131529 - 0.0019902 * SAT$$

Furthermore, using the median SAT score of 1040, we can expect a change of:

$$\frac{\partial I}{\partial \omega} = 2.131529 - 0.0019902 * 1040 = 0.061721$$

Considering that income and parental income are both measured in dollars, this means that for every parental income increase of \$1000, the average participant with a median SAT score can “expect” a corresponding increase in income by \$61.721. However, since the participant income variable was generated by summing annual income from four separate years, this increase in participant income is therefore distributed among the four years, i.e. the total income of the four years experiences a \$61.721 gain, or \$15.43025/year. Even in 1976 or 1979 dollars, this increase

is almost irrelevant when considering annual income. Although this relationship is statistically significant, the benefits of having wealthy parents are financially insignificant. Interestingly enough, while holding all other variables constant, the parental income coefficient indicates that having wealthy parents indeed leads to higher incomes, while the interaction coefficient is so slight that realistically, though statistically significant, it barely contributes to future earnings.

Next, still retaining SAT scores as the measure of potential academic performance, we add the parental education variable to examine the joint relationship. To run this regression we start by modeling total income with 0/1 dummies for the parental income levels of 1 (low) and 2 (medium) and SAT scores without the interaction effect:

$$I = \beta_0 + \beta_1 SAT + \beta_2 p_1 + \beta_3 p_2 + \varepsilon$$

In this model,  $p_1$  is the 0/1 dummy for having a parental education level of 1 while  $p_2$  is the 0/1 dummy for having a parental education level of 2. In this regression (See Regression 4, Table 1), only the SAT term is statistically significant at the 0.05 level with the following relationship:

$$\frac{\partial I}{\partial SAT} = \beta_1 = -12.34915$$

Within this model, a 100 point increase on the SAT corresponds to a decline in four-year income of \$1,234.915, a 2.9% fall from the median four-year income, hardly a financially significant change. Additionally, we see evidence that, by itself, parental education has no impact on future earnings.

Using the (n-1) 0/1 dummy model, we can set up the interaction effects of SAT scores and parental education level on total participant income (See Regression 5, Table 1). Here we see that the SAT term and both terms involving poorly educated parents are statistically significant. Once again, we take the partial derivative of income with respect to SAT scores:

$$\frac{\partial I}{\partial SAT} = \beta_1 + \beta_4 p_1 = -19.29838 + 36.5517 p_1$$

However, since  $p_1=1$  because we are looking at those with poorly educated parents:

$$\frac{\partial I}{\partial SAT} = 17.25332$$

This means that participants with poorly educated parents saw a positive relationship between income and SAT scores, with a 100 point increase on the SAT resulting in an income rise of \$1,725.332; considering the median income of \$43,400 for the children of poorly educated parents, this is just a 4.0% rise in four-year income.

Next, from the interaction term we see:

$$\frac{\partial I}{\partial p_1} = \beta_2 + \beta_4 SAT = -36362.32 + 36.5517 * SAT$$

Using the median SAT score for those with poorly educated parents, this becomes:

$$\frac{\partial I}{\partial p_1} = \beta_2 + \beta_4 SAT = -36362.32 + 36.5517 * 980 = -541.654$$

Although the units of this relationship are rather abstract, we care more about the magnitude of the result. Clearly, the average student with poorly educated parents who scores an average SAT score is at a tremendous disadvantage when it comes to income. At the same time however, such students can expect to see incomes rise with increasing SAT scores in accordance with:

$$\frac{\partial I}{\partial SAT \partial p_1} = \frac{\partial I}{\partial p_1 \partial SAT} = 36.5517$$

Given that, within this interaction effect relationship, the constant is statistically significant, we can additionally look at the impact of SAT scores and parental education on those with highly educated parents. Because we are using the (n-1) dummy model, the interaction effect between SAT scores and highly educated parents is rolled into the constant and SAT

coefficient. Here we are allowed to measure this relationship with a single partial derivative by setting  $p_1=p_2=0$ :

$$\frac{\partial I}{\partial SAT} = -19.29838$$

From these new relationships, we can thereby conclude that participants with poorly educated parents can improve their income by increasing their SAT scores while participants with highly educated parents see their incomes suffer with improved SAT scores. For a 100 point increase on the SAT, students with highly educated parents can expect to see a corresponding decrease of \$1,929.838 in income over the four years. This represents roughly a 4.8% loss out of the median four-year income of \$40,051.50 in this subgroup.

While these numbers are certainly statistically significant, they are, like the interaction effect of parental income and SAT scores on participant income, relatively financially insignificant. Although it is indeed interesting that the children of poorly educated parents saw increased incomes with improved SAT scores, possibly because those motivated or intelligent enough to improve their scores were incredibly driven individuals, the fact that the children of well-educated parents who experienced a negative relationship between SAT scores and income is seemingly inexplicable. At the same time however, the respective income gains/losses are each minute actual per-annum income changes. In relation to annual income, therefore, neither of these results in substantial financial relationships, and, indeed, even minor fluctuations in annual income (the  $\epsilon$  value in the model) can most likely override such a relationship in real-world circumstances. We therefore here conclude that, as a measure of background, level of parental education, in combination with participant SAT scores, does not correlate with participant income, i.e. that there is no noteworthy interaction effect of SAT scores and parental education on income.

Finally, we again change the measure of participant background and use the socioeconomic variable (See Regression 6, Table 1). Socioeconomic status is an ordered categorical variable with 1 denoting lower class; 2 indicating middle class; and 3 indicating upper class. As in the previous set of regressions, it is first necessary to generate 0/1 dummies for each socioeconomic status before running the regression. In running the regression, only the SAT coefficient and constant are statistically significant at the 0.05 level. It is therefore permissible to take the derivative of income with respect to SAT scores, yielding:

$$\frac{\partial I}{\partial SAT} = \beta_1 = -13.05819$$

It is entirely unsurprising that this value looks very similar to that of both SAT partial derivatives derived from the simple linear models of income based on SAT scores and parental income, and SAT scores and parental education. Since parental education and income are themselves components of the socioeconomic variable, one would expect the SAT partial derivatives, equaling the SAT coefficients, to be similar, as indeed they are. Thus far, ignoring the interaction effect, socioeconomic status does not affect participant income at a statistically significant level. In addition, as previously noted, increased SAT scores correlate with decreased income: here a 100 point increase on the SAT leads to a \$1,305.819 decline in four-year income, a 3.1% loss over the median four-year income.

Lastly, defining the interaction effect model of SAT scores and socioeconomic status is equivalently modeled similarly (See Regression 7, Table 1). Like in the previous regression series, the middle grouping variable, and its interaction effect with SAT scores, is not statistically significant at the 0.05 level, leaving statistically significant relationships between participants of lower and upper class with the interaction effect. Likewise, we can again take the partial derivative of income with respect to SAT scores:

$$\frac{\partial I}{\partial SAT} = \beta_1 + \beta_4 p_1 = -23.48172 + 63.07958 SES_1$$

Since we are dealing with  $SES_1=1$  for those of lower class backgrounds we can rewrite this as:

$$\frac{\partial I}{\partial SAT} = 39.59786$$

For every 100 point increase on the SAT, lower class students gain \$3,959.786 on their four-year income, a 9.3% gain over their median four-year income of \$42,700.

Next, by looking at the interaction term for participants from lower/class backgrounds:

$$\frac{\partial I}{\partial SES_1} = \beta_2 + \beta_4 SAT = -60541.1 + 63.07958 * SAT$$

Using the median SAT score for those from lower class backgrounds, this becomes:

$$\frac{\partial I}{\partial SES_1} = -60541.1 + 63.07958 * 965 = 330.6947$$

Again, it is the magnitude of this result we care about, regardless of the units. Within this relationship, we see that participants of lower class backgrounds can expect to see increases to income with improved SAT performance according to the mixed partial derivatives:

$$\frac{\partial I}{\partial SAT \partial SES_1} = \frac{\partial I}{\partial SES_1 \partial SAT} = 63.07958$$

Since the constant in this regression is statistically significant, we can take the partial derivative with respect to SAT scores and let  $SES_1=SES_2=0$  in order to explore the interaction effect between SAT scores and those of the upper socioeconomic class:

$$\frac{\partial I}{\partial SAT} = -23.48172$$

For upper/class participants, every 100/point increase on the SAT corresponds to a loss of \$2,348.172 from their median four-year income of \$41,450, a 5.7 % loss. Still, despite these changes being of the largest magnitude seen thus far, it is still important to remember that the

effects on income are spread out over four years, so even 9.3% is a marginal increase considering the timing.

### Regression Summary

In the simplest regression exploring the relationship between SAT scores and participant income, SAT scores were deemed to have a statistically significant negative impact, albeit at a financially trivial level. By regressing income, SAT scores, and parental income, we saw that only the SAT term affected income at a statistically significant level, again in a negative relationship, although again in a quantifiably small manner. When regressing between the same variables with the interaction effect, the parental income and interaction terms were statistically significant; specifically that having wealthy parents positively impacts future income, although only marginally upon later consideration.

In the regression between income, SAT scores, and parental education, the SAT term was again statistically significant in a negative relationship with income, although it barely affected participant income on a substantial financial level. When adding the interaction effect, students with poorly educated parents experienced increased incomes with improved SAT scores, although children with decreasingly well educated parents saw a drop in earnings.

Changing the participant background variable to socioeconomic standing, we saw that, overall, improved SAT scores led to a decrease in income. Within the interaction effect, students of low socioeconomic standing experienced statistically significant incomes gains of the largest magnitude seen in this study, although when considering the timing factor such gains seem less significant. Lower/class participants additionally saw increased earnings with improved SAT scores, while their upper/class peers saw income fall with improved SAT scores.

In every instance of statistically significant analysis, participants of privileged backgrounds were at a disadvantage at earning income; such students earned saw decreased earnings with improved SAT scores while their less privileged peers experienced gains with improved scores. Even broadly speaking, SAT scores correlated extremely poorly with participant income. Lastly, and perhaps most importantly, none of these findings agree with those presented in previous literature, i.e. privilege did not positively correlate with income or academic performance, as is widely accepted within the existing economic literature. Indeed, in our study, privilege did not seem to correlate at all with income, and participants across all socioeconomic classifications earned, on average, about the same.

#### **IV: Conclusion**

In short, although we had initially hypothesized that increased affluence and privilege would correlate with improved academic performance and increased income, our data analysis provides substantial evidence to reject our initial thoughts.

Moving towards a more formal examination of the data via the regressions, the results overall indicate a trivial financial impact due to students performance and participant background. The numerous insignificant financial findings of this study are at odds with all of the studies previously cited that found a positive correlation linking academic performance and income to socioeconomic status and other wealth considerations.

In each model, regression analysis indicated that there was no overall statistically significant relationship between participant background alone and income, without the interaction effect, suggesting that on average all participants earned the same income across possible background distributions; yet in controlling for background via the interaction effect there was a statistically significant difference between the effects of SAT scores in participant

income as noted above: those of lesser means saw improved incomes with increased SAT scores, while those of greater means experienced the opposite. Each of these findings disagrees with our initial hypotheses.

As shown within the regression analyses, both the standard linear model and interaction models resulted in inconclusive and trivial findings, i.e. in concert, performance and background did not impact income in a substantial manner. Both sets of results disagree with the existing literature.

In particular, the findings contradict Hamnett et al (2007) and Carneiro (2007), who found that children of increasingly wealthy families are more likely to perform well academically and go on to earn higher incomes. While Dobbie and Fryer (2011) cite “the problem of chronic underachievement [of less privileged children]”, here this is not entirely true; children of all backgrounds perform nearly the same on average when controlling for educational performance.

These conclusions are rather difficult to explain, especially given the current widespread conception that children of the affluent have an automatic advantage from birth. Today, discussions of income inequality as a result of socioeconomic stratification are quite common, yet the data here suggest the opposite. Bearing in mind that all participants included within the scope of this study graduated college in four years, perhaps these outcomes are not so difficult to explain after all.

One possible explanation is that the participants of lesser means, having demonstrated some level of academic dedication by graduating college within four years, were imbued with superior motivation and drive to succeed. Maybe these participants developed the best work ethic and as a result were rewarded with superior future earnings.

Indeed, the demonstrated increasing income with SAT scores, while controlling for participant background, makes for an excellent example. With fewer resources to devote towards preparing for the SAT, such driven yet less affluent participants studied hard to score well, and maybe carried such work ethic on to college and beyond, thus accounting for better earnings. Furthermore, considering that among participants belonging to the lower socioeconomic class, 96% (n=73) had low earning parents; with everything to gain, it is easy to see why these students could have been the best motivated and as such achieved superior incomes.

This trend fits into the larger historical narrative of the time, in which, following the closure of the Second World War, the ranks of the American middle class grew tremendously, with many profiting from increasingly upward socioeconomic mobility, partially driven by expanded access to higher education for the masses. The upward mobility clearly present within this dataset, evident from the children of the less-than-affluent going on to earn the same incomes as the children of the affluent, contradicts the notions of today, where common perception, and indeed numerous studies, have attested to the stagnation of social mobility and stratification. Above all else, analysis of the data supports the conclusion that trends in American education, income, and social mobility clearly change over time. It may seem a bit clichéd to say, but in one sense the findings support, at least for the participants of NLSC-72, that the “American Dream” of financial success through hard work and dedication was very much alive and well at the time.

At the same time, however, the limitations of the data herein drawn from the NLSC-72 must be noted in the context of this study. After data preparation, only 900 participants remained, an extremely small sample size considering the massive scope of NLSC-72 and the millions of college graduates who graduated from high school in 1972. It is not difficult to imagine that the

findings of this study are heavily dependent upon the 900 participants. Out of such an enormous population, 900 completely different participants could have led to equally different results.

In addition, the majority of the data was of the categorical format, and even though it was always “ordered” and our methods were well adapted to such a format, a dataset with more continuous variables would have been much more useful in terms of finding relevant relationships. In this sense, the data contained within NLSC-72 was not particularly well suited to our chosen methods of statistical analysis. Additionally, because the categories were themselves incredibly broad, it is quite possible that much of the nuanced relationships and interactions could have been missed by the analyses. To this end, we are essentially using the data for a means other than it was intended, and so the data is therefore naturally less than perfectly suited to our aims.

## **V: Reference List**

Carneiro, P. (2008). Equality of Opportunity and Educational Achievement in Portugal. *Portuguese Economic Journal*, 7(1): 17-41.

Das, J., Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, & Venkatesh Sundararaman. (2013). "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics*, 5(2): 29-57.

Dobbie, W.."( Roland Fryer. (2011). Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children’s Zone. *American Economic Journal: Applied Economics*, 3: 158-187.

Hamnett, C., Mark Ramsden, & Tim Butler. (2007). Social Background, Ethnicity, School Composition and Educational Attainment in East London. *Urban Studies Journal Foundation*, 44(7): 1255-1280.

Løken, K.V., Magne Mogstad, & Matthew Wiswall. (2012). What Linear Estimators Miss: The Effects of Family Income on Child Outcomes. *American Economic Journal*, 4(2): 1-35.

U.S. Dept. of Education, Center for Education Statistics. NATIONAL LONGITUDINAL STUDY OF THE CLASS OF 1972. ICPSR version. Chicago, IL: National Opinion Research

## VI: Regression Table

VARIABLES	(1) total income	(2) total income	(3) total income	(4) total income	(5) total income	(6) total income	(7) total income
SAT	-13.21** (5.797)	-13.65** (5.962)	10.56 (11.39)	-12.35** (5.908)	-19.30** (9.571)	-13.06** (5.977)	-23.48*** (7.878)
parents_income		0.0474 (0.148)	2.132** (0.850)				
SATxParents_income			-0.00199** (0.000799)				
p_educ1				810.1 (2,815)	-36,362** (16,027)		
p_educ2				4,626* (2,463)	9,948 (14,632)		
SATxParent_educ1					36.55** (15.23)		
SATxParent_educ2					-5,505 (13.56)		
ses_dummy1						589.9 (4,260)	-60,541*** (21,696)
ses_dummy2						48.61 (2,295)	-15,782 (13,506)
SATxSES1							63.08*** (22.06)
SATxSES2							14.93 (12.83)
Constant	61,548*** (6,160)	61,412*** (6,178)	36,683*** (11,686)	58,610*** (6,690)	66,197*** (10,598)	61,334*** (6,618)	72,613*** (8,639)
Observations	762	762	762	762	762	762	762
R-squared	0.007	0.007	0.015	0.012	0.023	0.007	0.018
Standard errors in parentheses							

\* p<0.01, \*\* p<0.05, \* p<0.1