Final Project **Darius Mehri**

Part 1. Background and Research Question

One of the more interesting questions in current sociological debates is how capital is used by individuals in society to reproduce their class position. One of the important theorists in the field today is Bourdieu who has posited a theory that individuals use various types of capital, such as economic, social and cultural capital, in a social field to increase their "life chances" to get ahead. An important aspect of Bourdieu's theory is the cognitive turn in understanding how capital is used to produce and reproduce class positions in society. Bourdieu expresses this theoretical concept through class the "habitus" which are the "predispositions of perceptions, thoughts and actions" of individuals in society (Bourdieu 1984). Understanding the habitus is particularly important since it influences how individuals use economic and social capital to get ahead in a social space, such as in a school, university and later in life when they enter the job market. The study of children is particularly important since the main principles behind Bourdieu's theory are that the habitus is shaped by one's education.

The research question for this paper will involve understanding the potential for building economic capital to increase "life chances" as related to future plans, social economic status and locus of control. The control variables in the analysis are gender and race. Math test scores are chosen as the response variable that measures the potential for economic capital since it is assumed that achievement in math opens the possibility for entering higher ranked colleges and careers in higher paying jobs, such as in the hard sciences. Social economic status and future plans are important for understanding how a child's current economic status and future plans are associated with high math test scores. The locus of control is important in understanding how cognition, i.e. the habitus, influences math scores and the ability to achieve economic capital.

Another aspect of investigating whether a child has the potential to build economic capital to improve life chances is disruption in education. This is measured through a logistic regression where controlling for race and gender, a child has been held back a grade. Other variables in the analysis include future plans, locus of control and social economic status. It is assumed that if a child is held back, his life chances will be dramatically reduced due to his or her inability to obtain economic capital.

Hypothesis 1a
High standardized math test scores will be associated with higher social economic status.

Hypothesis 1b
Children from races that carry a high social stigma will not experience a similar increase in standardized math test scores with increasing SES compared to white or Asian students.

The influence of lower social economic status on children's test scores is an indicator that the family's current economic capital has an influence on a child's ability to obtain future

economic capital. The measures of economic status will be based on parent's income and education levels (see below). Hypothesis 1b will be tested by interacting SES with race.

Hypothesis 2a
Children with higher standardized math test scores will have a greater locus of control and greater sense of self.

Hypothesis 2b
Children's higher standardized math test scores will be highly associated with future plans.

Children who do better in school must have greater self-esteem and feel as if they have greater control over their lives. The measurements include such things as feeling a sense of worth and the feeling that making plans in life leads to greater success. It is also expected that students with higher test scores will tend to want to achieve higher education to obtain the economic capital to improve life chances. This association will increase linearly with students future plans, for instance, children who plan on going to college will have higher test scores than those who do not.

Hypothesis 3a
The odds of children being held back will decrease linearly with increasing SES.

Hypothesis 3b
The odds of children being held back will be greater among races that carry a high stigma, such as African Americans, Hispanic or American Indians.

Hypothesis 4
The odds of children being held back will be greater for those children with a low locus of control, low self concept lower motivation to attend higher education after high school.

It is expected that children that are held back will have a greater odds of being from a lower SES and will not be highly motivated to succeed. Being held back will also be reflected in a greater odds of having a lower locus of control. All of these variables will contribute to a lower ability to obtain economic capital to improve life chances.

Part 2. Methods: Subjects and Variables

To test my hypothesis, I am using the National Education Longitudinal Study of trend data about critical transitions of 2000 8^{th} grade students. Its objective is to provide trend data about transitions experienced by students as they progress from elementary school through high school and into postsecondary education or the work force. The base year, used in this paper, is from a national sample of eighth graders along with their parents, teachers and school administrators. A subset of variables from this dataset was used for this project. I chose the math standardized test scores to measure the potential for children to build economic capital to succeed for the linear regression model. Being held back is the second response variable I use for the logistic regression model. I included a

number of explanatory variables including SES and variables to measure the locus of control and self concept. I also included other measures in an attempt to distinguish the most significant contribution to test scores. Here is a complete list of variables:

bymath: Math standardized test scores as continuous variables on a scale from 20.27 to 80.14. The mean is 51.5 and the standard deviation is 10.5

heldback: A binary variable on whether a student has been back by the 8th grade, 0 is not held back and 1 is held back.

male: dummy variable, 0 if female and 1 if male

byses: Social economic status measured as a continuous variable from -2.894 to 1.854 with a mean of zero and standard deviation of 0.8. It includes mother's and father's occupation and educational levels and total family income

race: Categorical variable, 1 – *api* (Asian Pacific Islander), 2- *hispanic*, 3- *African-American* (not Hispanic) 4 – *white* (not Hispanic) and 5 – *amindian* (Native American). Dummy variables were created for each category except for *white* which was the reference category for the analysis.

futureplans: Post secondary school plans are a categorical variable. Range is from 1 – won't finish high school, 2- graduate high school, 3- go to vocational school, 4 – attend college, 5- will graduate college, 6- attend higher level school after college. Dummy variables were created for each category and except for the reference category which was *won't finish high school.* For the logistic regression, due to the sparseness of those who are held back and who won't finish high school, I combined all of the children who plan on post-secondary school education in one variable called `fppostsec`

bylocus1: A continuous variable that measures the locus of control and ranges from -3 to 1.52 with a mean of zero and standard deviation of 0.7. It is based on a Likert scale of 1 – strongly agree, 2- agree, 3- disagree and 4- strongly disagree. The continuous scale is standardized and constructed from an answer to these questions:

I don't have enough control over the direction my life is taking

In my life, good luck is more important than hard work for success

Every time I try to get ahead, something or somebody stops me

My plans hardly ever work out, so planning only makes me unhappy

When I make plans, I am almost certain I can make them work

Chance and luck are very important for what happens in my life

bycncpt1: A continuous variable that ranges from -3.61 to 1.25 with a mean of zero, a standard deviation of 0.71 and is based on the same standardized Liker scale as bylocus1 but students answer these questions:

I feel good about myself

I feel I am a person of worth, the equal of other people

I am able to do things as well as most other people

On the whole, I am satisfied with myself

I certainly feel useless at times

At times I think I am no good at all

I feel I do not have much to be proud of

Part 3. Methods: Statistical analysis

To answer Hypothesis 1 and, I will create a linear regression model by regressing `bymath` on `byses`, `male`, `race`, `bylocus1`, `bycncpt1` and `futureplans`. I include race in the model to measure the effect not only of race on scores, but if the interaction between race and social economic status. For the logistic regression model, I will regress `heldback` on `byses`, `male`, `race`, `bylocus1`, `bycncpt1` and `futureplans`.

Part 4. Results: Descriptive Analyses

Two thousand observations were part of the dataset. Missing data points (i.e. people who did not answer questions properly for some of the questions) ranged from 15 to 45 per variable were not included in the analysis.

Linear Regression

Table 1: Continuous Variables: SES, Locus of Control and Self-Concept

Variable	Obs	Mean	Std. Dev.	Min	Max
byses	1773	-.0192594	.7912452	-2.894	1.854
bylocus1	1773	.039498	.6900774	-2.61	1.37
bycncpt1	1773	.0374563	.7151259	-3.45	1.12

Table 2: Gender

sex of respondent	Freq.	Percent	Cum.
male	863	48.67	48.67
female	910	51.33	100.00
Total	1,773	100.00	

Table 3: Race

r^s race/ethnic background	Freq.	Percent	Cum.
api	106	5.98	5.98
hispanic	225	12.69	18.67
black,non-hispanic	163	9.19	27.86
white,non-hispanic	1,215	68.53	96.39
american indian	64	3.61	100.00
Total	1,773	100.00	

Table 4: Post Secondary School Plans

post-secondary education plans	Freq.	Percent	Cum.
won^t finish h.s	11	0.62	0.62
will finish h.s	162	9.14	9.76
voc,trd,bus aftr h.s	134	7.56	17.32
will attend college	216	12.18	29.50
will finish college	776	43.77	73.27
higher sch aftr coll	474	26.73	100.00
Total	1,773	100.00	

Logistic Regression

Table 5: Held Back and Race

ever held back a grade in school	r^s race/ethnic background api	hispanic	black,non	white,non	american	Total
no	95	179	122	1,060	48	1,504
yes	11	46	41	155	16	269
Total	106	225	163	1,215	64	1,773

Table 6: Held Back and Gender

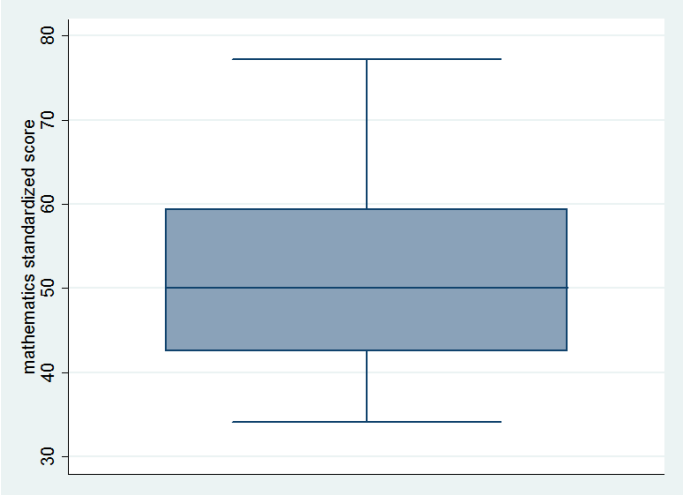
ever held back a grade in school	sex of respondent male	female	Total
no	712	792	1,504
yes	151	118	269
Total	863	910	1,773

Table 7: Held Back and Future Plans

ever held back a grade in school	post-secondary education plans won^t fin	will fini	voc,trd,b	will atte	will fini	higher sc	Total
no	7	104	96	170	699	428	1,504
yes	4	58	38	46	77	46	269
Total	11	162	134	216	776	474	1,773

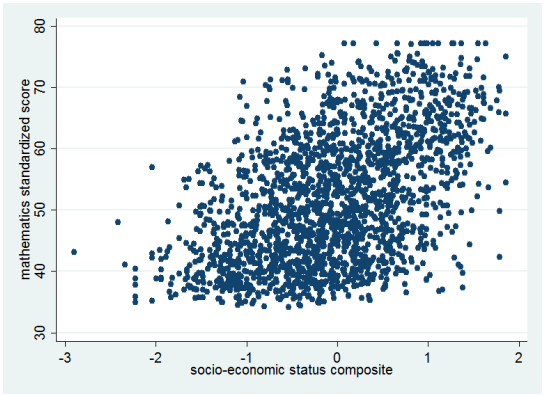
As can be seen from Table 6, the table is not sparse so there will not be a sampling size problem for held back and gender. Table 5 and 7 however are sparse. For being held back and future plans (Table 7), I combined the all students who plan on post-secondary school education into one variable called `fppostsec`. For Held Back and Race, I did not combine the variables since the combination of races is not possible.

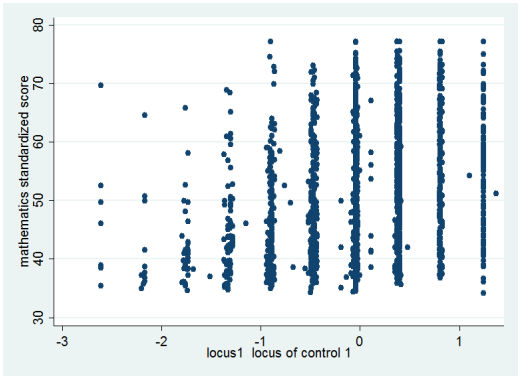
Box Plot of Response Variable – Standardized Math Scores:

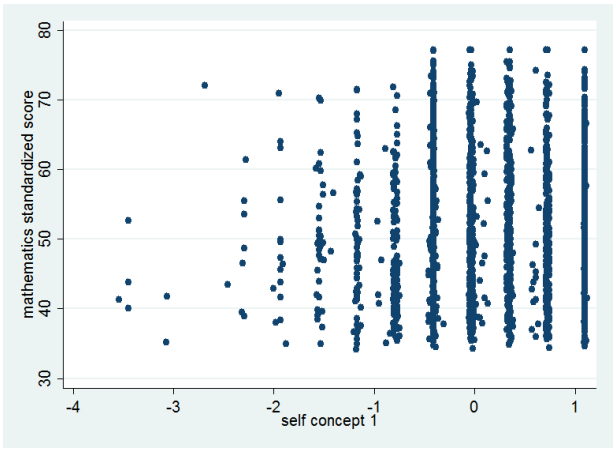


As can be seen from the plot, the response variable is not highly skewed and so the variable will not be transformed.

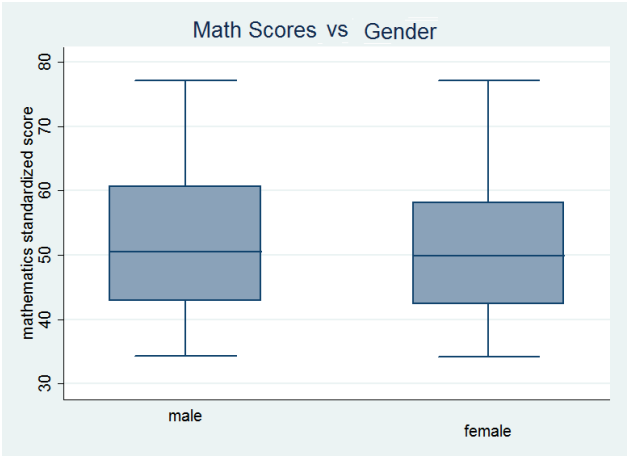
Scatter Plots of Response Variable with Continuous Explanatory Variables

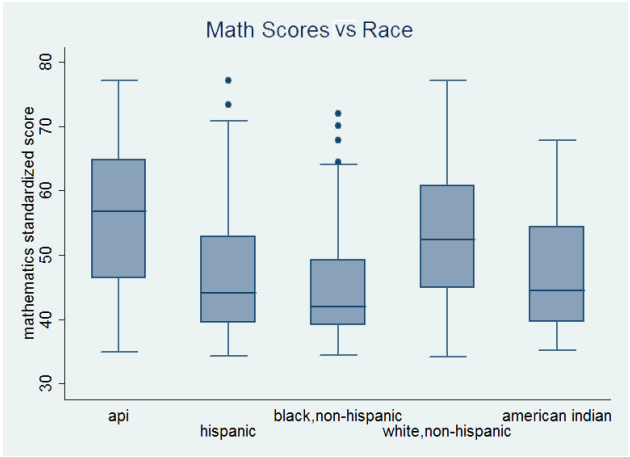


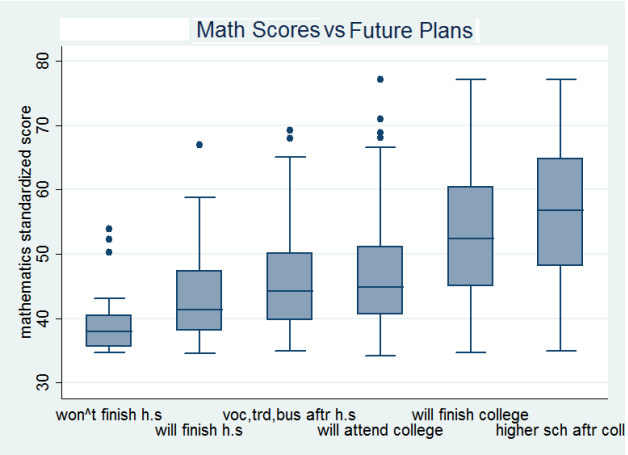




Box Plots of Response Variable with Categorical Variables







I don't see severe skewness in any of the variables so they will not be log transformed. There are some outliers in the variables as displayed in the plots but they do not seem to be severe so they will not be removed.

Part 5. Results: Regression analyses

Part a) Fit the Regression model

I ran two models for both the linear regression and logistic regression models, one with the response variable and the explanatory variables and the other with interactions. I used backward elimination to eliminate explanatory variables with $p>0.05$ and I checked for changes in R squared to ensure that it did not change significantly after a variable was dropped. I added higher order polynomial terms for continuous variables that were not initially significant to test whether there was a non-linear relationship between the response and explanatory variable.

Part b) Checking Collinearity

I tested collinearity using the Stata variance inflation factor `vif` command (see Appendix A). If the variance inflation factor was lower than 10, collinearity is not a problem among the explanatory variables in the model. In the models I tested, I did not find any collinearity so variables were not removed or combined.

Part c) Assessing Linearity

I assessed linearity by graphing the predicted response variable against the continuous variables – locus of control, social economic status and self concept. All three of the relationships are linear and the variables therefore do not need to be transformed. See the graphs in Appendix B.

Part d) Assessing Constant Variance and Normality of Residuals

I used two methods to assess the linear regression assumptions: Partial Plots (see Appendix A) to show constant variance and linearity and a comparison of the models with and without the Stata `robust` command to test for changes in model results. Using both of these procedures, I did not find heteroskedasticity in the model i.e. there were no

changes in the regression coefficients or R^2 when running the model using robust. I used the kernel density estimate of the residuals to test for normality: the residuals seemed relatively normal so the assumption of normality is valid (see Appendix B). A model comparison of the significance of categorical variables was conducted using the Stata `testparm` command before including them in the model (see the Table and Appendix A for results).

Part e) Estimates and Discussion

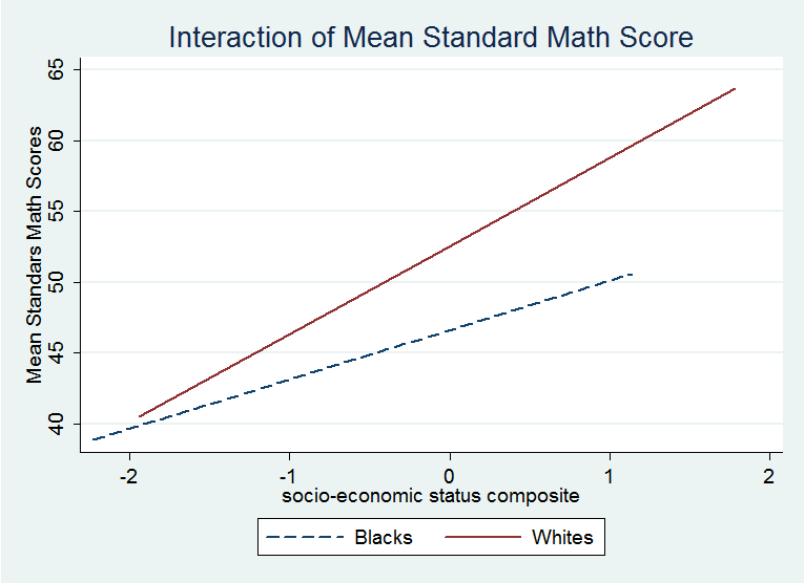
Linear Regression Model

As can be seen from Model 1 in Table 1 (see next page), the continuous variables that were significant are SES and locus of control. Self concept was not significant and was dropped from the model. I added higher order polynomial terms to test whether self-concept was significant and they were not and so they were dropped from the model. Analyzing the Pearson Correlations the most significant positive influence on the estimated math scores for continuous variables is SES (r=0.28) and locus of control (r=0.14). For categorical variables it is future plans on obtaining a college degree and obtaining a degree beyond college (r=0.20 and 0.23 respectively). The largest negative influence on test scores is children who are African-American (r=-0.18)

Controlling for the other explanatory variables, a 1 unit increase in SES results in an approximate 3.75 point increase in the estimated mean standardized math score and a 1 unit increase in locus of control results in a 1.84 increase in the estimated mean standardized math score. Race is also significantly associated with math scores (F=22.76), on average when controlling for the other explanatory variables, African-Americans, Native Americans and Hispanics have a 5.5, 3.7 and 2.1 estimated mean standard score lower than whites respectively. Asian Pacific Islanders, on the other hand, have a 2.4 higher estimated mean standard score compared to whites. Men score 1.3 points higher compared to females. The most significant effect on the mean standard math scores was future plans (F=39.17). Comparing children who had plans to finish high school to those who did not, there was no significant difference in mean standard scores. The differences rose when children stated they planned on attending a vocational school or college, mean standard scores on average are higher by 2.27 and 2.37 respectively compared to those who would not finish high school. For students who planed on finishing college or who had plans on post college degrees, the mean standard math scores on average increase by 6.3 and 8.05 points respectively.

Interactions

I interacted race and future plans on SES, and locus of control. The only variables that turned out to be significant were the interactions between African-American and SES, and Native Americans attending college. The interaction between African-American and SES reveals that on average, controlling for other explanatory variables, if SES increases by one unit, African-American children have an estimated mean standard score rate of increase of 1.8 *less* compared to whites. Here is a graph of the predicted mean standard score and SES:



The other interaction that was significant was American Indians who plan to attend college having a -6.4 mean standard score less than whites who plan to attend college.

Logistic Regression

The logistic regression (see results in Table 2 below) shows that being held back a grade is positively correlated with being male and African-American but negatively correlated with students who plan on attending a higher level education school upon graduating from high school. Being held back is also negatively correlated with an increase in social economic status and locus of control. Controlling for other variables, a one unit increase in SES reduces the odds of being held back by 47% and a one unit increase in locus of control reduces the odds by 25%. Controlling for other variables, attending education after high school reduces the odds of being held back by 37%. Compared to those students not planning on a post-high school education. Controlling for all other variables, males have a 59% greater odds of being held back compared to females. All variables were significant at the 5% level. Those variables that were not significant at the 5% level such as self-concept and some of the race variables were dropped from the model. I interacted SES on the race variables and none of the interactions were significant. After discovering that self concept was not significant, I added quadratic and cubic polynomial variables but they were not significant and were dropped from the model.

Table 2: Logistic Regression Results

R^2=0.088, LR Chi2 (4) = 229, p<0.0001

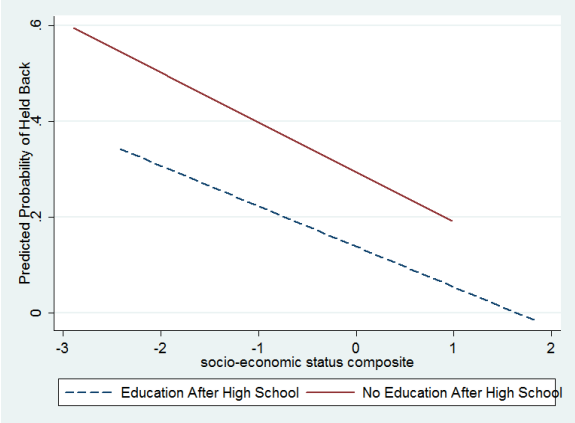
Held Back	**Odds Ratio**	**Std. Err.**	**P>|t|**	**[95% Conf. Interval]**	
SES	*0.53*	0.01	<0.001	0.59	0.88
Locus of Control	0.75	0.07	0.003	0.62	0.91
Male	*1.59*	0.22	0.001	1.30	2.26

African-American	1.77	0.36	0.006	1.18	2.64
fppostsec	*0.63*	0.12	<0.001	0.43	0.93

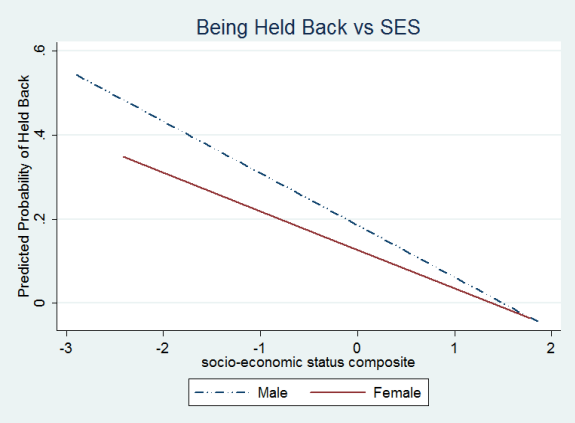
Graphs of Predicted Probabilities

Graphing the predicted probabilities with the continuous variables revealed some interesting trends. Graph 2 shows that the probability of being held back decreases with social economic status and that the differences in the probability remain constant if the person is planning on attending education after high school or not. Graph 3, however, shows that for students of lower SES males have a much larger probability of being held back compared to females but the differences diminish with increasing SES. Graph 4 shows a similar trend comparing African-American students with non-African-American students. The probability of being held back for African-American students with a low SES is much higher compared with non-African-American students but the gap decreases to zero with increasing SES. Graph 5 shows that there is no change in the gap in the probability of being held back if the student is African-American compared to non-African-American with increasing locus of control.

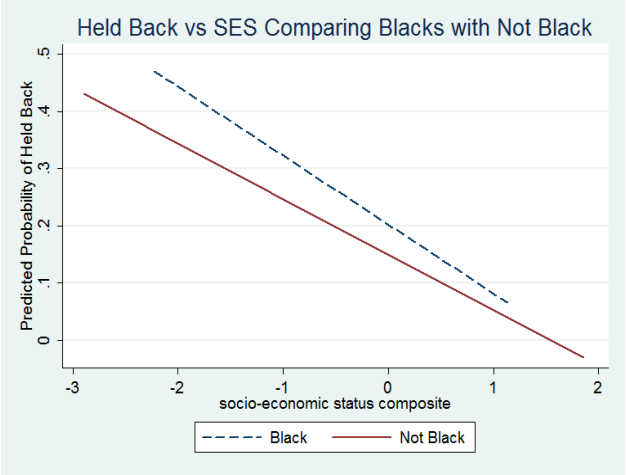
Graph 2



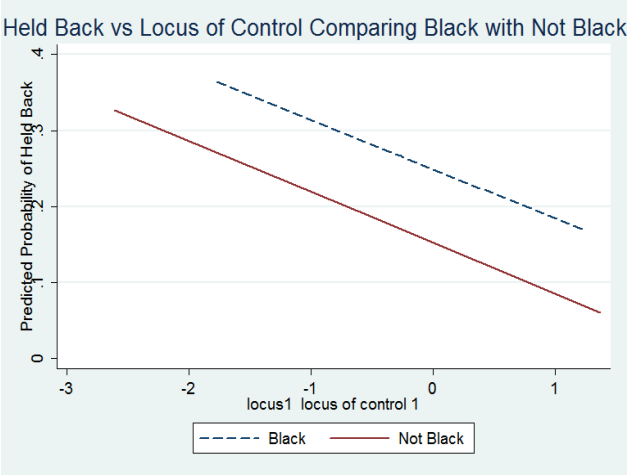
Graph 3



Graph 4



Graph 5



Part 6. Discussion

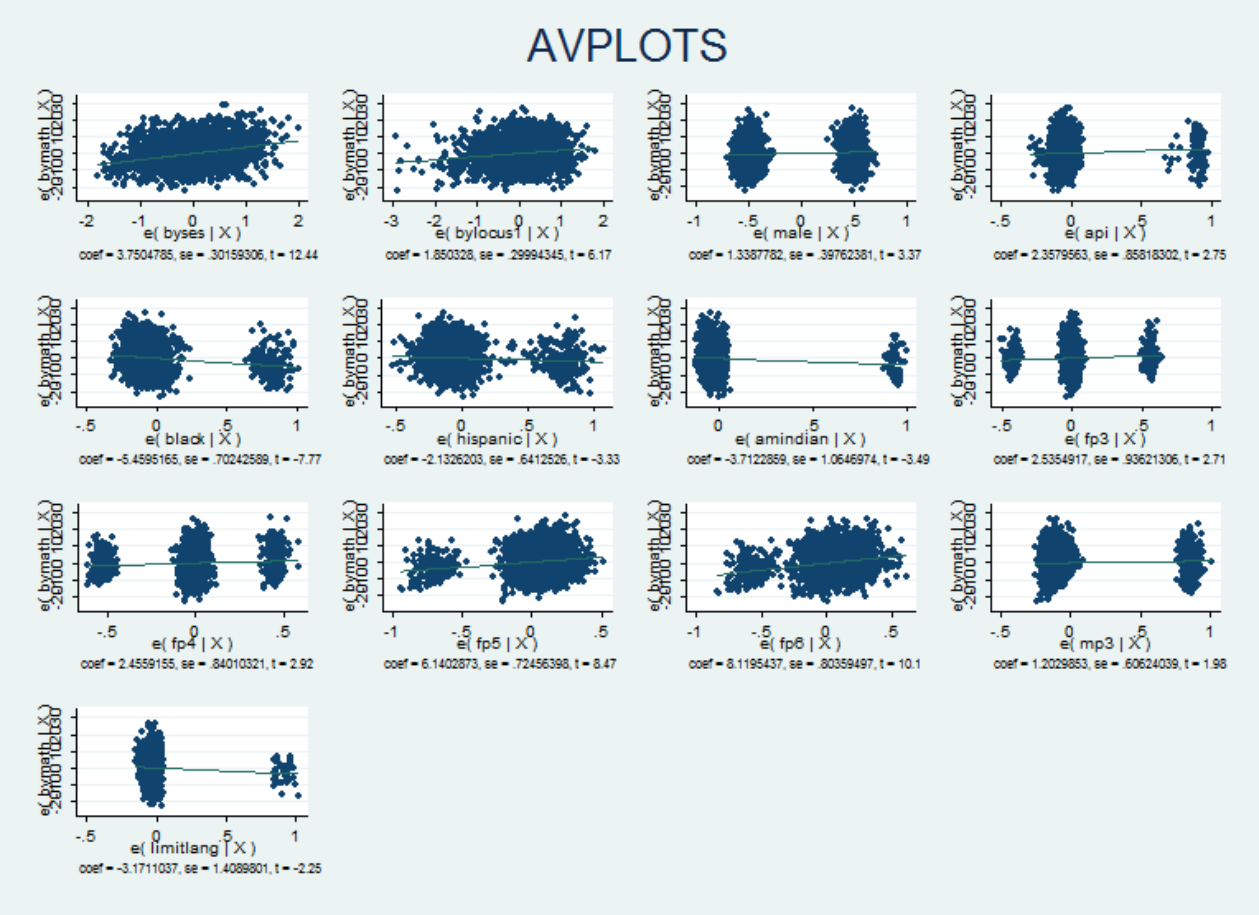
The linear regression results show strong support for Hypothesis 1a - Children with lower test scores will be from families with lower social economic statuses. Race plays a very large role and there is a ranking with Asian Pacific Islanders having the highest scores, whites second, Hispanics third, Native Americans fourth and African-Americans last. African-Americans also have a reduced rate of increase in estimated standard math scores compared to whites when SES increases. Children with higher motivation as reflected in future plans have a significantly higher estimated math score, having a stronger locus of control increases estimated test scores slightly but self concept has no influence on test scores. This is partial support for Hypothesis 2.

The logistic regression results show support for Hypothesis 3a – the odds of children being held back decrease linearly with increasing SES. The results however show partial support for Hypothesis 3b since the odds of being held back are significant only for African-Americans. The results also show a partial support for Hypothesis 4. Being held back increases with locus of control but there is no association with being held back and self concept. There is however a very high association between being held back and future plans – those students who plan on attending a vocational school or college after high school have a consistently lower probability of being held back with increasing SES. There is however a high association with being held back and gender – for lower SES there is a much higher probability for males to be held back than females but this gap shrinks to zero with increasing SES.

These results support the Bourdieun social theory that children who are expected to succeed come from background with higher economic capital as reflected in higher math scores and lower probabilities of being held back as associated with higher SES, but also have the cultural and social capital to succeed as reflected in higher motivation to achieve higher education and a greater locus of control.

APPENDIX A

Partial Plots (Check for constant variance)



Testing for significance of categorical variables

Future plans
. testparm fp1 fp2 fp3 fp4 fp5 fp6

(1) fp3 = 0
(2) fp4 = 0
(3) fp5 = 0
(4) fp6 = 0

F(4, 1773) = 39.17
Prob > F = 0.0000

Race
testparm white African-American api hispanic amindian

(1) African-American = 0

(2) api = 0
(3) hispanic = 0
(4) amindian_fp5 = 0

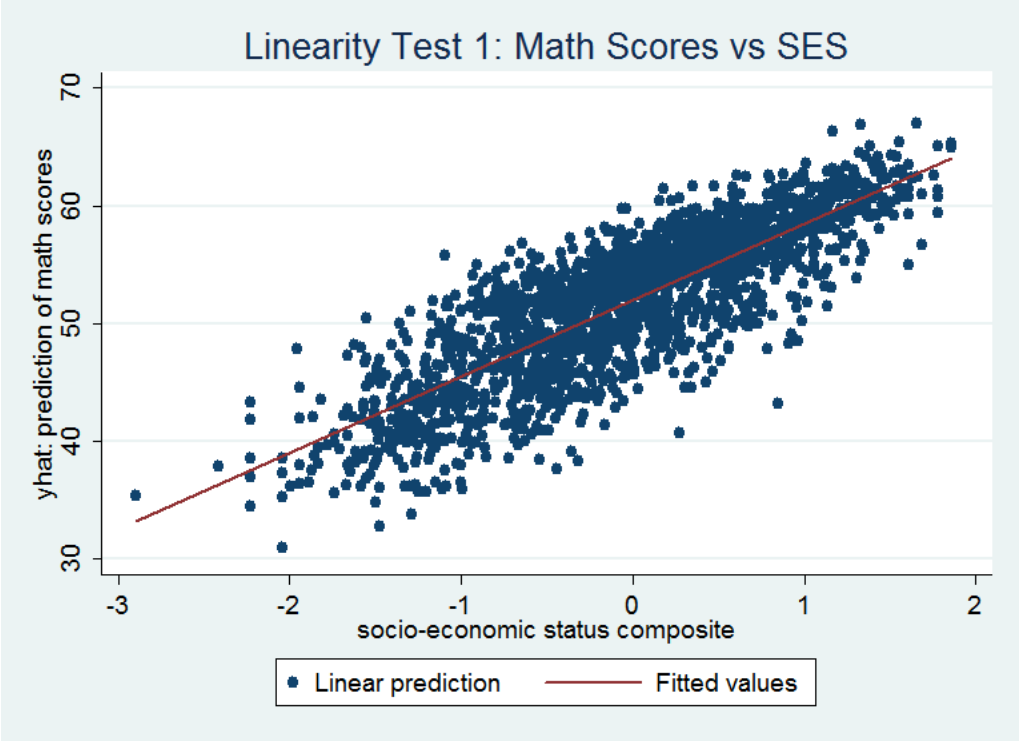
F(4, 1773) =   22.76
Prob > F =   0.0000

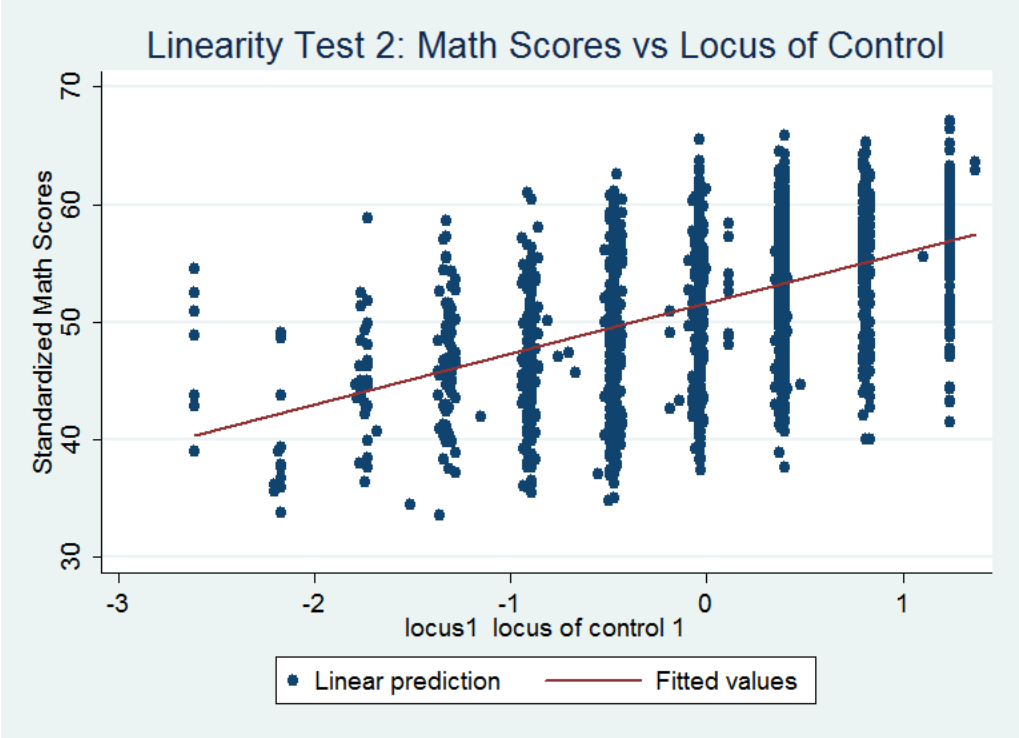
Collinearity Test

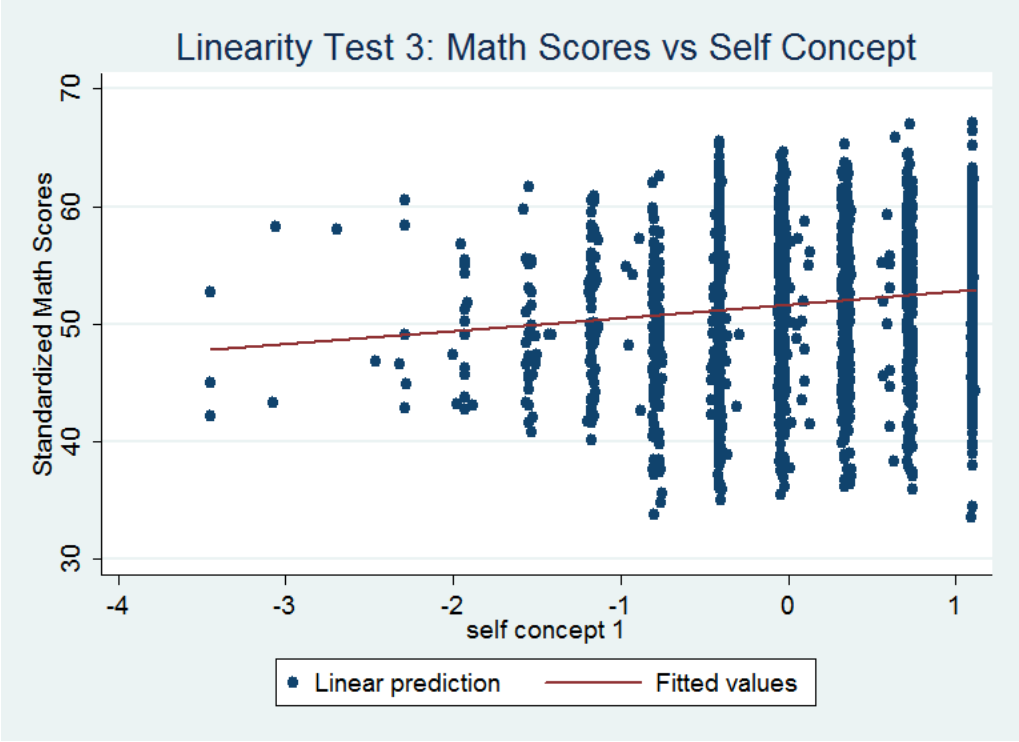
```
. vif
```

Variable	VIF	1/VIF
fp5	3.57	0.279772
fp6	3.46	0.289164
fp4	2.02	0.495744
fp3	1.67	0.599513
byses	1.58	0.631515
black_byses	1.40	0.712180
black	1.30	0.769933
hispanic	1.14	0.874633
bylocus1	1.10	0.906470
api	1.04	0.963437
amindian_fp5	1.03	0.969288
male	1.02	0.984213
Mean VIF	1.69	

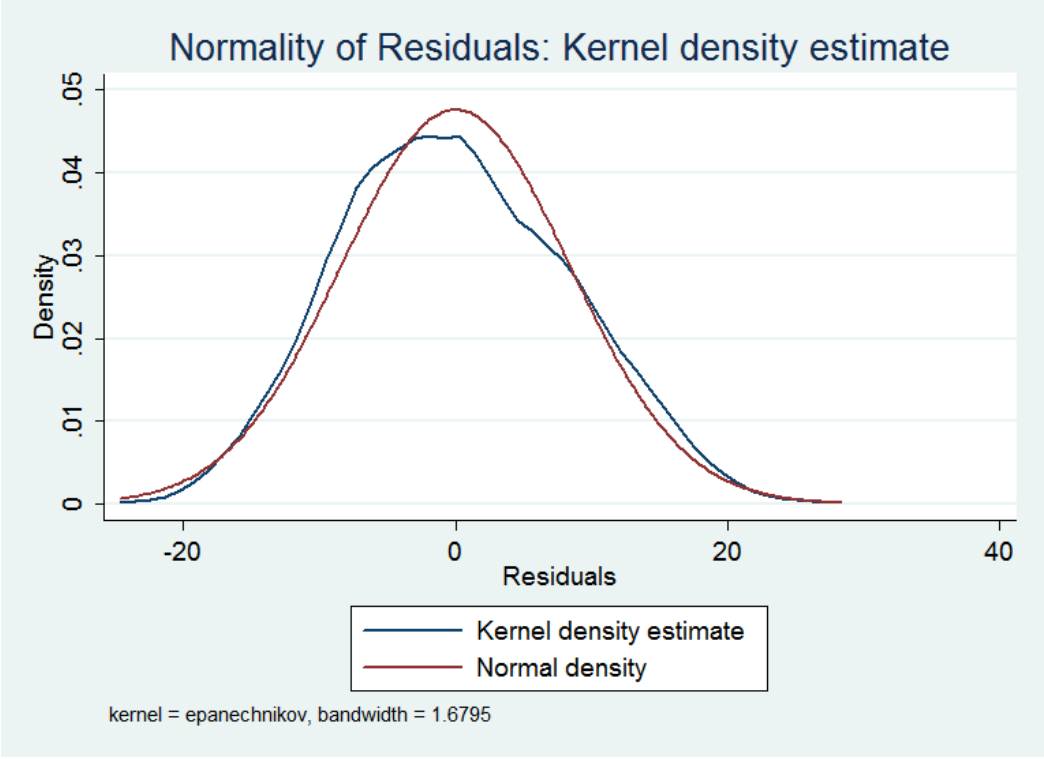
APPENDIX B


Linearity Test 1: Math Scores vs SES
yhat: prediction of math scores
30
40
50
60
70
-3
-2
-1
0
1
2
socio-economic status composite
Linear prediction
Fitted values


Linearity Test 2: Math Scores vs Locus of Control
Standardized Math Scores
30
40
50
60
70
-3
-2
-1
0
1
locus1 locus of control 1
Linear prediction
Fitted values



Normality of Residuals



APPENDIX C

DO FILE

```
/// nels data
use "C:\Users\Darius\Documents\My Work\Berkeley Classes Spring
2009\Education Stats Class\final paper\nels_data.dta", clear

keep bypsepln bycncpt1 bys12 by2xmstd byses bys31a bylocus1 bys74


// BEING HELD BACK
replace bys74 = . if bys74 > 3
rename bys74 heldback
//change no of 1 to 0, yes of 2 to 1
recode heldback 1=0
recode heldback 2=1


// GENDER VARIABLE, MALE =1, FEMALE =2
replace bys12 = . if bys12==8
rename bys12 gender
tabulate gender, generate(d)
rename d1 male
rename d2 female


// SES
replace byses = . if byses > 90

//LOCUS OF CONTROL
replace bylocus1 = . if bylocus1 > 99.98

//SELF CONCEPT
replace bycncpt1 = . if bycncpt1 > 90
//square and cube for checking significance of higher order terms
generate sqr_bycncpt1 = bycncpt1*bycncpt1
generate cube_bycncpt1 = bycncpt1*bycncpt1*bycncpt1


// FUTURE PLANS
replace bypsepln = . if bypsepln > 90
rename bypsepln futureplans
//creates categorical variables for future plans
tabulate futureplans, generate(fp)

//RACE AND ETHNICITY
//drop if bys31a > 5
```

```
replace bys31a = . if bys31a == 6
replace bys31a = . if bys31a == 7
replace bys31a = . if bys31a == 8
rename bys31a race
tabulate race, generate(r)
rename r1 api
rename r2 hispanic
rename r3 African-American
rename r4 white
rename r5 amindian


// BY2XMSTD MATHEMATICS STANDARDIZED SCORE
//drop if by2xmstd > 97
replace by2xmstd = . if by2xmstd > 99
rename by2xmstd bymath



//INTERACTIONS
// MALE AND AFRICAN-AMERICAN
generate male_African-American = male*African-American
// MALE AND LOCUS OF CONTROL
generate male_bylocus1 = male*bylocus1
//RACE AND BYLOCUS
generate white_bylocus1 = white*bylocus1
generate African-American_bylocus1 = African-American*bylocus1
generate hispanic_bylocus1 = hispanic*bylocus1
generate api_bylocus1 = api*bylocus1
generate amindian_bylocus1 = amindian*bylocus1
//RACE SES
generate white_byses = white*byses
generate African-American_byses = African-American*byses
generate hispanic_byses = hispanic*byses
generate api_byses = api*byses
generate amindian_byses = amindian*byses
// RACE FUTURE PLANS
generate white_fp5 = white*fp5
generate African-American_fp5 = African-American*fp5
generate hispanic_fp5 = hispanic*fp5
generate api_fp5 = api*fp5
generate amindian_fp5 = amindian*fp5


//FINAL MODEL
#delimit ;
regress bymath byses bylocus1 male api African-American hispanic fp3
fp4 fp5 fp6 African-American_byses amindian_fp5, robust;
#delimit cr


predict yhat, xb
predict r, resid


//COMBINE VARIABLES
generate nonwhite = African-American + hispanic + amindian
```

```
//future plans, fphschool = have high school or lower education,
fpcollege = college or higher
generate fphschool = fp1 + fp2
generate fppostsec = fp3 + fp4 + fp5 +fp6

generate fppostsec_African-American = fppostsec*African-American
generate fppostsec_hispanic = fppostsec*hispanic
generate fppostsec_amindian = fppostsec*amindian

//LOGISTIC REGRESSION
//final model
logistic heldback bymath byses male fppostsec

predict predprob, pr
```

References:

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Cambridge, MA: Harvard University Press.