What Matters Most: Teasing Out the Predictive Power of Early Performance on College Readiness

Benjamin Howe and Avis Atkins

Harvard Graduate School of Education

December 17, 2012

Abstract

A number of empirical research studies have established the predictive power of early academic performance on college readiness. Yet, very few research considers the academic curriculum variation around this predictive relationship, and even less research has considered how this relationship differs based on the type of high school. Thus, researchers conducted a series of regression analyses to understand how a high school's academic curriculum influences (ACI) the relationship between early academic performance and college readiness and additional regression analysis to explore how this relationship differs by school type. The multiple regression model revealed a three-way interaction between eighth grade math standardized test scores, ACI, and school type in predicting a student's math SAT scores. Analysis showed that there was an estimated -0.24 point difference in the effect of eighth grade standardized math scores on math SAT scores for every one unit difference in ACI for private school students, however this relationship did not hold for students in public schools. This means that ACI in high school for private schools can help improve a student's college readiness, as measured by math SAT scores, but this effect does not seem to hold true in public schools.

keywords: college readiness, early academic performance, SAT scores

What Matters Most: Teasing Out the Predictive Power of

Early Performance on College Readiness

A number of empirical research studies have established the predictive power of early academic performance on college readiness. Yet, very few research considers the academic curriculum variation around this predictive relationship, and even less research has considered how this relationship differs based on the type of high school.

Thus, we explored two main research questions. First, is the established relationship between early academic performance and college readiness different based on a school's academic curriculum intensity (ACI)? Secondly, does this assumed association between early academic performance, academic curriculum intensity, and college readiness, matter less for students in private schools as compared to students in public schools?

We were looking at important contributing factors for college readiness. If we can find important links to college readiness that possibly predict this, then it would be important for future policy makers so that they can help address the most critical issues related to improving college readiness.

In this model, (See *Figure X.1*) it was hypothesized that the rigor of a school's curriculum will have differential effects on the relationship between eighth grade math standardized test scores and SAT math scores. ACI allows a student a somewhat second chance on achieving a high math SAT score. In other words, ACI reduces the total overall effect that eighth grade math standardized math scores has on predicting a student's math SAT scores. Furthermore, it was hypothesized that the effect of the relationship between eighth grade standardized test scores, SAT math scores, and academic curriculum intensity would be smaller in size for private school students.

The purpose of the current study was to describe the predictive power of early academic performance on college readiness by examining how this relationship differs based on the rigor of the school and also how the type of school influences this relationship. The paper is structured as follows. The methods section presents the survey used to evaluate the research question as well as a description of the design, sample, and measures used in the study. The results section includes a summary of the analysis performed, the model building strategy, and the final model selection. Finally, the discussion section interprets the findings and also includes implications, limitations, and suggestions for future research.

Methods

Design and Sample

Data on individual variables and other demographics came from the National Educational Longitudinal Study of 1988 (NELS: 88) administered by the U.S. Department of Education's National Center for Education Statistics (NCES) (NCES, 1989). In 1988, the NCES randomly selected approximately 1,000 public and private middle schools from a universe file of 40,000 schools. Approximately 24 eighth graders were randomly selected from each middle school. Asian and Latino students were oversubscribed in each school sample.

Selected students were surveyed in the spring of 1988. A sample of the surveyed respondents was then resurveyed through four follow-ups in 1990, 1992, 1994, and 2000. The questionnaire collected student-level characteristics like parents education, information on reading, social studies, and math achievement tests, and self-reports on alcohol use, smoking, and drug use. The original NELS panel of 1988-2000 consisted of approximately 12,100 students. For this study we examined a sub-sample of respondents (N=700) who took the math section of the SAT.

Measures

College readiness. College readiness was our main outcome measure. College readiness was measured using students' math SAT score. Math SAT score was an interval variable ranging from 230 to 800.

Early academic performance. Drawing on prior literature we included early academic performance as our main predictor of college readiness. Early academic performance was measured using eighth grade standardized test scores. Eighth grade standardized math test score was a continuous variable ranging from 35.69 to 77.2.

Academic curriculum intensity. A high school's academic curriculum intensity (ACI) and type of school were included as potential moderators in the model. A high school's academic curriculum intensity was included as a potential moderator of the relationship between eighth grade standardized math scores and math SAT scores. For the NELS:88, levels of academic intensity of a student's high school curriculum were determined based on a review of student high school transcripts. The methodology computes an intensity level that includes the highest level of curriculum across major subject areas. A high school's academic curriculum intensity was a discrete continuous variable ranging from 1 to 31.

School type. The type of high school, public versus private, was included as a potential moderator of the relationship between eighth grade standardized test scores, math SAT scores, and the school's academic curriculum intensity. High school type was a binary variable coded as zero and one indicating a public school and private school respectively.

Demographics and parent's educational background. Prior literature suggests that demographic and parents' educational background is a source of differences in college readiness. Although the aim of the study was not the role of demographic and parent-level characteristics, we included socioeconomic status (SES), gender, race, and parents education as controls. SES was a composite score of parents education and occupational prestige, household possessions, and total family income on a scale from 0 to 6, with 6 being the highest family SES level. Gender was the student's sex, which was a binary variable with zero indicating male and one indicating female. Race, a variable indicating the student's race, was a binary variable with one indicating non-white and zero indicating white.

Analysis

Before constructing and testing the model, it was first necessary to determine if there was, in fact, a relationship between eighth grade standardized math scores and SAT math scores. Data were analyzed using the Stata statistical program (StataCorp., 2011). The correlation matrix reported with the descriptive statistics in *Table M.2* shows that a statistically significant association between eighth grade standardized math scores and math SAT scores exists.

We also conducted a series of regression analyses to examine the relationships between the control variables in the model and the other criterion variables. Eighth grade standardized test scores were used to predict SAT math scores. In order to explore if the established relationship between early academic performance and college readiness differed based on a school's academic curriculum intensity, we ran a multiple regression analysis. Furthermore, to examine if the assumed association between early academic performance, academic curriculum intensity, and college readiness, mattered less for students in private schools as compared to students in public schools we ran another multiple regression analysis. We then added the product terms between math and academic curriculum intensity, high school type and eighth grade math scores, high school type and academic curriculum intensity, and the three way product term of high school type, eighth grade standardized math scores, and academic curriculum intensity. Finally, we added the control variables, SES, gender, parents education, and race into the regression equation to predict math SAT scores.

There was no need to eliminate any terms in our model, since there was a three-way interaction term between eighth grade standardized math scores, ACI, and school type, and all lower order terms were necessary for this interaction in running the multiple regression analysis.

Results

Univariate descriptives. Preliminary univariate descriptives revealed math SAT scores were unimodal and symmetric (M = 501.55, SD = 115.79) and scores ranged from 230 to 800. Eighth grade standardized math scores were unimodal, symmetric, and approximately normal (M = 57.55, SD = 9.61) and scores ranged from 35.69 to 77.20. ACI was unimodal and skewed left (M = 20.95, SD = 6.95) and scores ranged from 1 to 31. SES was unimodal and skewed left (M = 3.32, SD = 0.72) and ranged from 0.96 to 5.30 (See *Table M.1*).

Bivariate descriptives. Furthermore, bivariate descriptives showed there appeared to be a moderately strong positive linear association with our main interest of a relationship between eighth grade standardized math scores versus math SAT scores as indicated by looking at the scatter plot. Eighth grade standardized math scores explained 81 percent of the variability in math SAT scores and it was statistically significant at the $\alpha = 0.05$ level ($r = 0.81$, $p < 0.001$). There was also a statistically significant effect of eighth grade standardized math scores on math SAT scores ($t_{obs} = 36.57$, $df = 698$, $p < 0.001$), such that there was an estimated difference of 9.77 (95% CI: [9.24, 10.30]) points in mean standardized math test scores corresponding to a one point differences in eighth grade standardized math scores. Furthermore, there did not appear to be any outliers in the scatter plot.

Model building. Since eighth grade standardized math scores was our predictor of interest, the first regression model fit (Model M.1 in *Table M.4*) is a simple linear regression with eighth grade standardized math scores as the only predictor. We identified ACI and school type as predictors of interest as well, so they enter our second regression model fit (Model M.2 in *Table M.4*). The next regression model looks at all our predictors and their interactions, regardless of statistical significance in the previous model (Model M.3 in *Table M.4*). In this model, we were interested in looking at the possible moderating effects of ACI on eighth grade standardized math scores, school type possible moderating effects on ACI and eighth grade standardized math scores, and school type possible moderating effects on ACI moderation of eighth grade standardized math scores. Our final model (Model M.4 in *Table M.4*) includes socioeconomic status, student gender, race, and parents education, because they could act as a likely confounder of our relationship and were necessary to use as control variables in studying the relationship between math SAT scores and eighth grade standardized math scores, ACI, and school type. Since the three-way interaction between eighth grade standardized math scores, ACI, and school type was significant at the $\alpha = 0.05$ level ($b = -1.41$, $t = -2.60$, $df = 688$, $p = 0.01$;) no trimming of the model was necessary and everything else remained in the model (regardless of its statistical significance) for statistical control. All models were compared using nested F-tests and regression coefficient t-tests as appropriate.

Final model assumptions. Linearity appeared to hold for the multiple regression model and was determined by examining the plot of studentized residuals of math SAT scores versus eighth grade standardized math scores, the plot of studentized residuals of math SAT scores versus eighth grade standardized math scores by school type, and the plot of studentized residuals of math SAT scores versus ACI. All the residual points were randomly scattered and not in some sort of pattern such as a u-shape, indicating possibly curvilinearity, which would have lead to concern of linearity being violated.

Homoscedasticity also appeared to hold for the multiple regression model and was determined by examining the plot of studentized residuals of math SAT scores versus the predicted math SAT scores and the plot of studentized residuals squared of math SAT scores versus the predicted math SAT scores. All the residual points were randomly scattered and not in some sort of a pattern, such as a megaphone shape indicating heteroscedasticity, thus there was no concern of violation of homoscedasticity.

Lastly, normality appeared to hold for the multiple regression model and was determined by examining the histogram of the math SAT scores studentized residuals with a normality curve overlaid. Since the histogram was unimodal and symmetric, there was no concern about a violation of normality.

We also performed a sensitivity analysis since *Cook's D* revealed a possible influential point (Student 1495450, *Cook's D* = 0.03). However, the sensitivity analysis regression output with and without this possible influential point revealed no real difference in the R^2 (0.716 vs. 0.715), suggesting student 1495450 was not influential, and by not having any or many extreme outliers we were not worried about it violating normality as well.

Final model. Thus we were left with the generic equation $\widehat{SATm} = \hat{\beta}_0 + \hat{\beta}_1(Math8) + \hat{\beta}_2(SES) + \hat{\beta}_3(Gender) + \hat{\beta}_4(Race) + \hat{\beta}_5(Parent\ Education) + \hat{\beta}_6(ACI) + \hat{\beta}_7(School\ Type) + \hat{\beta}_8(Math8 \times ACI) + \hat{\beta}_9(School\ Type \times Math8) + \hat{\beta}_{10}(School\ Type \times ACI) + \hat{\beta}_{11}(School\ Type \times Math8 \times ACI)$. This turns out to equal Model M.4 in *Table M.4* or $\widehat{SATm} = -53.12 + 7.83(Math8) + 13.09(SES) - 19.36(Gender) + 0.34(Race) + 15.01(Parent\ Education) + 1.13(ACI) - 295.07(School\ Type) + 0.03(Math8 \times ACI) + 5.22(School\ Type \times Math8) + 15.71(School\ Type \times$

ACI) - 0.27(*School Type x Math8 x ACI*) after the model building and is reflected in *Table M.5*. The final model explained 71.59 percent of the variability in math SAT scores and it was statistically significant at the $\alpha = 0.05$ level ($F(11, 688) = 157.57$, $p < 0.001$).

Statistical results of the final model. There was a statistically significant effect of ACI and eighth grade standardized math scores on math SAT scores in private schools, controlling for family SES, student gender, race, and parents education ($b = -0.24$, $t = -2.47$, $df = 688$, $p = 0.014$). There was also a statistically significant effect of eighth grade standardized math scores on math SAT scores in private schools, controlling for family SES, student gender, race, and parents education ($b = 7.56$, $t = 8.61$, $df = 688$, $p < 0.001$). There is an estimated difference of 7.56 (95% CI: [5.83, 9.28]) points in math SAT scores corresponding to a one point difference in eighth grade standardized math scores for private school students. Overall, this means there was an estimated -0.24 (95% CI: [-0.43, -0.05]) point difference in the effect of eighth grade standardized math scores on math SAT scores for every one unit difference in ACI for private school students. Substantially this means that an ACI of thirty-one could reduce eighth grade standardized math scores explanatory effect on math SAT by nearly 7.44 points versus an ACI of one and only 0.24 points. Through looking at *Figure X.2*, we see that higher eighth grade math scores and higher ACI corresponds to higher math SAT scores. However, *Figure X.2* reveals that a student in private school with low eighth grade standardized math scores has a better chance of achieving higher math SAT scores if there ACI is high and this effect begins to diminish as the students eighth grade standardized math scores become higher. At a certain eighth grade standardized math score, which seems to be approximately 70 through looking at *Figure X.2*, the relationship begins to switch and the high ACI begins to diminish their improvement on math SAT scores versus a student with a lower ACI.

However, there was not a statistically significant moderating effect of ACI on the relationship between eighth grade standardized math scores and math SAT scores for public schools, controlling for family SES, student gender, race, and parents education ($b = 0.03$, $t = 0.84$, $df = 688$, $p = 0.399$). But there was a statistically significant effect of eighth grade standardized math scores on math SAT scores, controlling for family SES, student gender, race, and parents education ($b = 7.83$, $t = 9.36$, $df = 688$, $p < 0.001$) for public schools. There is an estimated difference of 7.81 (95% CI: [6.18, 9.47]) points in math SAT scores corresponding to a one point difference in eighth grade standardized math scores for public school students. Through looking at *Figure X.2*, substantially this means that higher eighth grade math scores corresponds to higher math SAT scores, but ACI will not be substantially helpful in improving math SAT scores for public school students.

Discussion

Our model revealed that for all public school students, higher ACI lead to slightly higher SAT scores regardless of 8th grade math ability. However, for private school students that were lower performing in 8th grade, a higher ACI was associated with a higher SAT score compared with their similar ability peers in less intense ACI tracks. Thus, a high school's academic rigor did serve to possibly influence the relationship between early academic performance and later college readiness. Moreover, the relationship between early academic performance, college readiness, and ACI was different for students in private schools versus public schools.

The implications of our study showed that a student's eighth grade standardized math score is highly predictive of their success on their math SAT scores. We believe that the SAT can be used as a predictor of college readiness, so higher SAT scores would correlate to being more prepared for college. We also found that ACI could play a significant role in acting as another way to become more prepared for college and the math SAT during high school. This means that if a child struggled in the eighth grade on their math scores, by taking a more rigorous school load during high school, it improved their chances of achieving higher scores on the math SAT.

However, this effect was only true for private schools and not public schools. This means that if a student struggled in private school on their eighth grade math scores and then took a rigorous course load in high school, they could greatly increase their math SAT scores and improve their college readiness. But as the student had a higher eighth grade math score, their level of ACI became less important in improving their math SAT scores. This means at some point no matter how high the private schools student's ACI was, if there eighth grade math score was high enough, ACI would not play a role in improving the math SAT scores any higher. In fact at a certain eighth grade standardized math score, which seems to be approximately 70 through looking at *Figure X.2*, the relationship begins to switch and the high ACI begins to diminish their improvement on math SAT scores versus a student with a lower ACI. This could mean that if a student is smart enough in the eighth grade, as measured by their math scores, if they take too high of an academic load in high school it could be detrimental, such as the load is too heavy to actually learn all the material they are studying that helps prepare them for college. Yet, these are only speculations and future research should be done to study this phenomenon observed in our model.

We also found that ACI did not play a role in helping a public school student improve their math SAT scores regardless of their eighth grade standardized math scores. This is alarming because it could possibly indicate a major flaw in the public school system. It seems classes, no matter how hard they are, are not effective in improving a student's readiness for college. These students seemed to be bound and determined by whatever their standardized scores at a young age were and that would mostly determine their scores later on in age. Policy makers and school administrators should take serious note to this. This evidence suggest that they should take a close look at their curriculums and improve their classes, so that if a student struggles at a young age they can become academically better at a later age, versus being stuck in the same track of very little improvement. Future research should definitely be done to see if these speculations could be correct. It would be interesting to see what factors are really contributing to a public school student's success if the intensity of the academic curricula cannot improve their college readiness.

The limitations to our study are that while we have explained 71.59 percent of the variation in contributing factors to the outcome of a student's math SAT scores, which we believe could be used as a predictive measure of academic success, there is still about 30 percent of the variability left to be explained. The National Educational Longitudinal Study of 1988 looks at other factors such as grade point average, class rank, time spent on homework, whether a high school is urban or not, and the high school socioeconomic status that were not included in our model that could explain some of the other 30 percent of the variability in explaining a student's math SAT score. There could also be many other contributing factors that were not tested in this study as well.

Future research would need to look at these other variables and other variables yet to be discovered that could lead to explaining more of the variation in a student's math SAT scores. We should also address that while we used math SAT scores as a possible predictor of college readiness, other measures could be analyzed as well, such as grade point average. We only addressed math SAT scores and the SAT consists of two other sections, reading and writing, that should be analyzed as well in comparing a student's standardized reading and writing scores in the eighth grade. Some students may excel in other subjects and our study only looked at math, which could show that some subjects as being a valid measure of predicting a student's college readiness and others as not. It would also be important for future researchers to see what subjects are the best at improving a student for college readiness as various fields and careers require different background preparations.

The established association between early academic performance and college readiness is not a simple one. Based on this study there seems to be multiple factors influencing this relationship. If society's goal is to ensure that we understand what may predict a student's future readiness to go to college, then we need to understand how different factors interact to explain this general relationship. For now, it seems as if a school's academic rigor is only helpful for some students. We want to ensure that policies are helpful for all students.

References

StataCorp. (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

U.S. Department of Education, National Center for Education Statistics. (1989). National Education Longitudinal Study, 1988. Chicago, IL: National Opinion Research Center.

Table M.1

Descriptive Statistics for SAT Math Scores and Predictor Variables (Eighth Grade Standardized Math Scores, Family SES, Student Gender, Race, Parent Education, ACI and School Type), n=700

Variable	M	SD	Min	Max
Eighth Grade Standardized Math Scores	57.55	9.61	35.69	77.20
Family SES	3.32	0.72	0.96	5.30
ACI (1-31)	20.95	6.95	1	31
Student Gender (1=female; 0=male)	53.1% female; 46.9% male			
School Type (1=private; 0=public)	19.1% private; 80.9% public			
Race (1=non-white; 0=white)	27.1% non-white; 72.9% white			
Parent Education (1=at least one parent with a college degree; 0=no parent with a college degree)	49.3% college degree; 50.7% no college degree			

Table M.2

Pairwise Correlations, Partial Correlations and Semipartial R2 Values for Math SAT Scores with Predictor Variables (Eighth Grade Standardized Math Scores, Family SES, Student Gender, Race, Parent Education, ACI and School Type), n=700

Outcome	Predictors	Pairwise Correlation	Partial Correlation	Semipartial R^2
Math SAT Scores	Eighth Grade Standardized Math Scores	0.81***	0.74***	0.36***
	Family SES	0.41***	0.09*	0.002*
	Student Gender	-0.11**	-0.15***	0.01***
	Race	-0.07	0.01	<0.001
	Parent Education	0.36***	0.08*	0.002*
	ACI	0.50***	0.28***	0.02
	School Type	0.12***	-0.02	<0.001

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Table M.3

Pairwise Correlations and Tolerance for Model Predictor Variables (Eighth Grade Standardized Math Scores, Family SES, Student Gender, Race, Parent Education, ACI and School Type), n=700

	Student Gender	Family SES	Eighth Grade Math Scores	ACI	School Type	Race	Parent Education	Tolerance
Student Gender	1.00							0.98
Family SES	-0.11**	1.00						0.38
Eighth Grade Math Scores	-0.03	0.34***	1.00					0.76
ACI	0.04	0.24***	0.42***	1.00				0.78
School Type	-0.10*	0.22***	0.10**	0.17***	1.00			0.93
Race	0.01	-0.17***	-0.09*	0.05	-0.09*	1.00		0.94
Parent Education	-0.10**	0.77***	0.28***	0.18***	0.16***	-0.07	1.00	0.41

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Table M.4

Results of Fitting a Taxonomy of Multiple Regression Models for Student Math SAT Scores as a Function of Eight Grade Math Scores, High School Academic Curriculum Intensity, and School Type and Their Interactions, Controlling For Family SES, Student Gender, Race, and Parent's Education in a Random Sample of n = 700 American High School Students

	Parameter Estimate (*se*)			
	Model M.1	Model M.2	Model M.3	**Model M.4**
Intercept	-60.79*** (15.59)	-70.78*** (15.00)	-37.88 (47.26)	**-53.12 (46.58)**
Eight Grade Math Scores (Math8)	9.77*** (0.27)	8.81*** (0.28)	8.23*** (0.86)	**7.83*** (0.84)**
High School Academic Curriculum Intensity (ACI)		3.08*** (0.39)	1.01 (2.32)	**1.13 (2.25)**
School Type (1=private; 0=public)		5.15 (6.34)	-406.77** (148.77)	**-295.07* (143.93)**
(Math8) x (ACI)			0.04 (0.04)	**0.03 (0.04)**
(School Type) x (Math8)			7.18** (2.61)	**5.22* (2.53)**
(School Type) x (ACI)			20.10** (6.39)	**15.71* (6.18)**
(School Type) x (Math8) x (ACI)			-0.34** (0.11)	**-0.27** (0.11)**
SES				**13.09* (5.35)**
Student Gender (1 = female; 0 = male)				**-19.36*** (4.79)**
Race (0 = White; 1 = Not White)				**0.34 (5.47)**
Parents Education (0 = no parent with a college degree; 1 = at least one parent with a college degree)				**15.01* (7.36)**
Root MSE	67.85	65.00	64.74	**62.21**
R^2	0.66	0.69	0.69	**0.72**
Model *F*-test	13337.65***	507.32***	221.23***	**157.57*** **
(df_1, df_2)	(1, 698)	(3, 696)	(7, 692)	**(11, 688)**

*p<0.05, **p<0.01, ***p<0.001

Table M.5

Results for the Final Regression Model M.4 of Math SAT Scores on Eighth Grade Standardized Math Scores, Family SES, Student Gender, Race, Parent Education, ACI and School Type (n = 700)*

	Est.	*se*	*p*-value	95% C.I.	*Beta*
Intercept	-53.12	47.29	0.26	[-145.98, 39.73]	-
Eighth Grade Math Score	7.83	0.84	<0.001	[6.19, 9.48	0.65
Family SES	13.09	5.35	0.02	[2.58, 23.60]	0.08
Student Gender (1=female; 0=male)	-19.36	4.79	<0.001	[-28.76, -9.96]	-0.08
Race (1=non-white; 0=white)	0.34	5.48	0.95	[-10.42, 11.10]	0.001
Parent Education (1=at least one parent with a college degree; 0=no parent with a college degree)	15.01	7.36	0.04	[0.55, 29.47]	0.06
ACI (1-31)	1.13	2.25	0.62	[-3.28, 5.54]	0.07
School Type (1=private; 0=public)	-295.07	143.93	0.04	[-577.66, -12.48]	-1.00
(8th Grade Math Scores) x (ACI)	0.03	0.04	0.40	[-0.04, 0.11]	0.15
(School Type) x (8th Grade Math Scores)	5.22	2.53	0.04	[0.26, 10.18]	1.07
(School Type) x (ACI)	15.71	6.18	0.01	[3.57, 27.85]	1.30
(School Type) x (8th Grade Math Scores) x (ACI)	-0.27	0.11	0.01	[-0.48, -0.07]	-1.41
Root MSE	62.21				

*$R^2 = 0.72$; $F(11,688) = 157.57$, $p < .001$

Figure X.1. Hypothesized model.

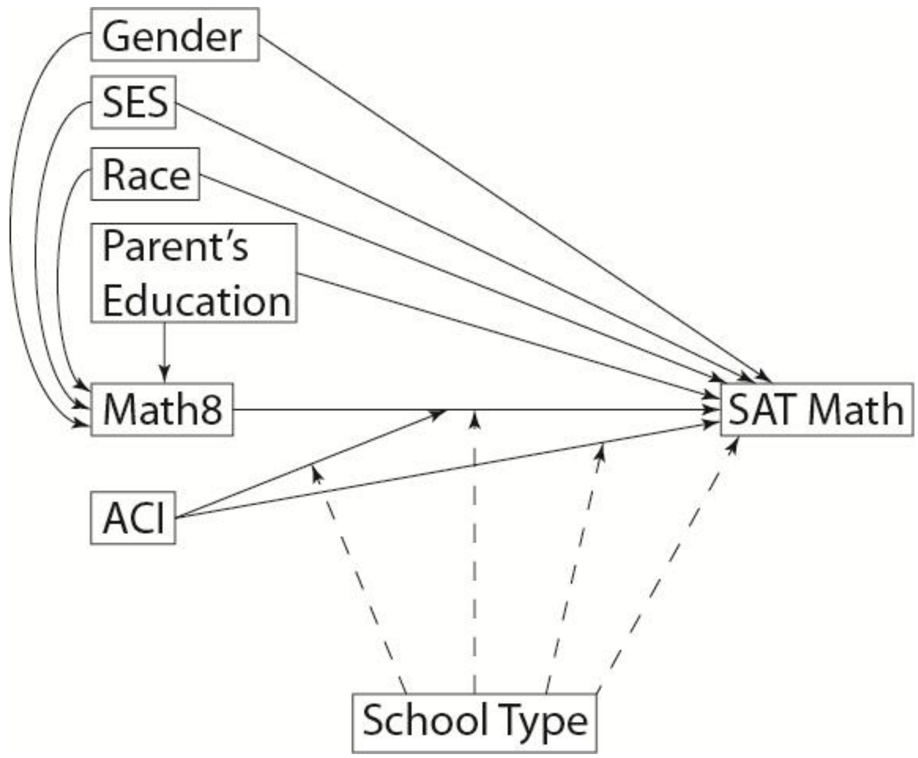
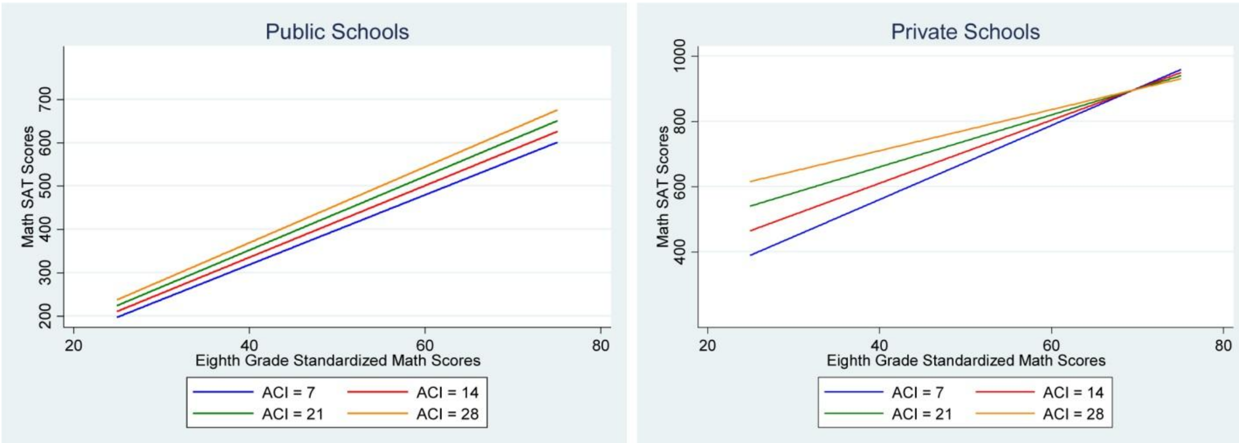


Figure X.2. Fitted values of eighth grade standardized math scores and ACI by school type as a function of math SAT scores, controlling for family SES, student gender, ace, and parents education.



Appendix

Stata Do File:

```
*****
*opens log file for results
log using "C:\Users\hgseuser\Desktop\FinalProject.smcl", replace
*open and loads data file
use "C:\Users\hgseuser\Desktop\FinalProject_NELS88.dta", clear
*****
*Univariate Descriptives
*gives you univariate descriptives for math SAT scores
tabstat satm, stat(mean sd range min p25 p50 p75 max iqr)
*gives you univariate descriptives for socioeconomic status
tabstat ses, stat(mean sd range min p25 p50 p75 max iqr)
*gives you univariate descriptives for eighth grade math standardized test scores
tabstat math8, stat(mean sd range min p25 p50 p75 max iqr)
*gives you univariate descriptives for High school academic curriculum intensity
tabstat aci, stat(mean sd range min p25 p50 p75 max iqr)
*gives you univariate descriptives for High school type (private vs. public)
tabstat hs_private, stat(mean sd range min p25 p50 p75 max iqr)
*gives you univariate descriptives for gender (male vs. female)
tabstat female, stat(mean sd range min p25 p50 p75 max iqr)
*gives you univariate descriptives for parent education
tabstat pared, stat(mean sd range min p25 p50 p75 max iqr)

*generates new variable of white or non-white
gen nwhite = race!=4
*labels new variable nwhite
label variable nwhite "nwhite"

*gives you univariate descriptives for race (white vs. non-white)
tabstat nwhite, stat(mean sd range min p25 p50 p75 max iqr)

*****
*Bivariate Descriptives
*gives you a histogram for math SAT scores
histogram satm, bin (19) ///
name(satmhist, replace) ///
saving("C:\Users\hgseuser\Desktop\satmhist.gph", replace)
*gives you a histogram for socioeconomic status
histogram ses, bin (25) ///
name(seshist, replace) ///
saving("C:\Users\hgseuser\Desktop\seshist.gph", replace)
*gives you a histogram for eighth grade math standardized test scores
histogram math8, bin (19) ///
```

```
name(math8hist, replace) ///
saving("C:\Users\hgseuser\Desktop\math8hist.gph", replace)
*gives you a histogram for High school academic curriculum intensity
histogram aci, bin (29) ///
name(acihist, replace) ///
saving("C:\Users\hgseuser\Desktop\acihist.gph", replace)
*gives you a histogram for High school academic curriculum intensity
histogram female, bin (3) ///
name(femalehist, replace) ///
saving("C:\Users\hgseuser\Desktop\femalehist.gph", replace)
*gives you a histogram for school type
histogram hs_private, bin (3) ///
name(hs_privatehist, replace) ///
saving("C:\Users\hgseuser\Desktop\hs_privatehist.gph", replace)
*gives you a histogram for race
histogram nwhite, bin (3) ///
name(nwhitehist, replace) ///
saving("C:\Users\hgseuser\Desktop\nwhitehist.gph", replace)
*gives you a histogram for parent education
histogram pared, bin (3) ///
name(parenteducationhist, replace) ///
saving("C:\Users\hgseuser\Desktop\parenteducationhist.gph", replace)

*produces a scatterplot of satm versus math8
scatter satm math8
name(satm vs math8, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsmath8scatter.gph", replace)
*produces a scatterplot of satm versus math8 with line of best fit
graph twoway (scatter satm math8) (lfit satm math8), ///
name(satmVmath8, replace) legend(off) ///
xtitle("Eighth Grade Standardized Math Scores") ytitle("Math SAT Scores")
saving("C:\Users\hgseuser\Desktop\math8Vsatmwline.gph", replace)
*produces a scatterplot of satm versus aci
scatter satm aci
name(satm vs aci, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsaciscatter.gph", replace)
*produces a scatterplot of satm versus ses
scatter satm ses
name(satm vs ses, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvssesscatter.gph", replace)
*produces a scatterplot of satm versus gender
scatter satm female
name(satm vs female, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsfemalescatter.gph", replace)
*produces a scatterplot of satm versus school type (public vs private)
scatter satm hs_private
```

```
name(satm vs school type, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsshooltypescatter.gph", replace)
*produces a scatterplot of math8 versus ses
scatter math8 ses
name(math8 vs ses type, replace) ///
saving("C:\Users\hgseuser\Desktop\math8vssesscatter.gph", replace)
*produces scatterplots of satm versus math8 by gender
scatter satm math8, by(female)
name(satm vs math8 by gender type, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsmath8bygenderscatter.gph", replace)


*produces scatterplots of satm versus math8 by hs_private
scatter satm math8, by(hs_private)
name(satm vs math8 by hs_private type, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsmath8byhs_privatescatter.gph", replace)
*produces scatterplots of satm versus math8 by high_ses & hs_private
scatter satm math8, by(high_ses hs_private)
name(satm vs math8 by high_ses & hs_private type, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsmath8byhigh_ses&hs_privatescatter.gph", replace)
*produces scatterplots of satm versus math8 by female & hs_private
scatter satm math8, by(female hs_private)
name(satm vs math8 by female & hs_private type, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsmath8byfemale&hs_privatescatter.gph", replace)
*produces scatterplots of satm versus math8 by aci & hs_private
scatter satm math8, by(aci hs_private)
name(satm vs math8 by aci & hs_private type, replace) ///
saving("C:\Users\hgseuser\Desktop\satmvsmath8byaci&hs_privatescatter.gph", replace)
*pairwise correlation of all predictor, control, and outcome variables
pwcorr satm female ses math8 aci hs_private nwhite pared, sig
*partial and semipartial correlations of SAT math scores with predictor and control variables
pcorr satm female ses math8 aci hs_private nwhite pared
*looks for collin command and click to install
findit collin
*collinearrity diagnostics for predictor and control variables
collin female ses math8 aci hs_private nwhite pared

*****
*Creates new variable math8xaci
generate math8xaci = math8*aci
*Labels new variable math8xaci
label  variable math8xaci "math8xaci"

*Creates new variable math8xaci
generate hs_privatexaci = hs_private*aci
*Labels new variable math8xaci
```

```
label  variable hs_privatexaci "hs_privatexaci"

*Creates new variable math8xaci
generate hs_privatexmath8 = hs_private*math8
*Labels new variable math8xaci
label  variable hs_privatexmath8 "hs_privatexmath8"

*Creates new variable math8xaci
generate hs_privatexfemale = hs_private*female
*Labels new variable math8xaci
label  variable hs_privatexfemale "hs_privatexfemale"

*Creates new variable math8xaci
generate hs_privatexmath8xaci = hs_private*math8*aci
*Labels new variable math8xaci
label  variable hs_privatexmath8xaci "hs_privatexmath8xaci"

REGRESSIONS - Model One
*runs a regression of SAT math scores on eighth grade math standardized test scores
regress satm math8
*runs a regression of SAT math scores on eighth grade math standardized test scores, high
school academic curriculum intensity, and status of high school
regress satm math8 aci hs_private
*runs a regression of SAT math scores on eighth grade math standardized test scores, high
school academic curriculum intensity, status of high school
*the product of standardized 8th grade math scores and high school academic curriculum
intensity, the product of high school type and eighth grade math scores
*the product of high school type and high school academic curriculum intensity, and the
interaction term of high school type, eighth grade math scores, and academic
*curriculum intensity
regress satm math8 aci hs_private math8xaci hs_privatexmath8 hs_privatexaci
hs_privatexmath8xaci
*runs a regression of SAT math scores on eighth grade math standardized test scores, high
school academic curriculum intensity, status of high school
*the product of standardized 8th grade math scores and high school academic curriculum
intensity, the product of high school type and eighth grade math scores
*the product of high school type and high school academic curriculum intensity, and the
interaction term of high school type, eighth grade math scores, and academic
*curriculum intensity controlling for ses, gender, race, and parents' education
regress satm math8 aci hs_private math8xaci hs_privatexmath8 hs_privatexaci
hs_privatexmath8xaci ses female nwhite pared

REGRESSIONS - Model Two
*runs a regression of SAT math scores on eighth grade math standardized test scores
regress satm math8
*runs a regression of SAT math scores on eighth grade math standardized test scores controlling
```

```
for ses, gender, race, and parents' education
regress satm math8 ses female nwhite pared
*runs a regression of SAT math scores on eighth grade math standardized test scores, high
school academic curriculum intensity, and status of high school
*controlling for ses, gender, race, and parents' education
regress satm math8 ses female nwhite pared aci hs_private
*runs a regression of SAT math scores on eighth grade math standardized test scores, high
school academic curriculum intensity, status of high school
*the product of standardized 8th grade math scores and high school academic curriculum
intensity, the product of high school type and eighth grade math scores
*the product of high school type and high school academic curriculum intensity, and the
interaction term of high school type, eighth grade math scores, and academic
*curriculum intensity
regress satm math8 ses female nwhite pared aci hs_private math8xaci hs_privatexmath8
hs_privatexaci hs_privatexmath8xaci
*runs a regression of SAT math scores on eighth grade math stadardized test scores, high school
academic curriculum intensity, status of high school
*the product of standardized 8th grade math scores and high school academic curriculum
intensity, the product of high school type and eighth grade math scores
*the product of high school type and high school academic curriculum intensity, and the
interaction term of high school type, eighth grade math scores, and academic
*curriculum intensity with beta weights
regress satm math8 ses female nwhite pared aci hs_private math8xaci hs_privatexmath8
hs_privatexaci hs_privatexmath8xaci, beta


*Graph for public schools*
graph twoway ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*7 + _b[hs_privatexaci]*(0*7) + (_b[math8] + _b[math8xaci]*7 +
_b[hs_privatexmath8]*0 + _b[hs_privatexmath8xaci]*(0*7))*x, ///
range(25 75) clcolor(blue)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*14 + _b[hs_privatexaci]*(0*14) + (_b[math8] + _b[math8xaci]*14
+ _b[hs_privatexmath8]*0 + _b[hs_privatexmath8xaci]*(0*14))*x, ///
range(25 75) clcolor(red)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*21 + _b[hs_privatexaci]*(0*21) + (_b[math8] + _b[math8xaci]*21
+ _b[hs_privatexmath8]*0 + _b[hs_privatexmath8xaci]*(0*21))*x, ///
range(25 75) clcolor(green)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*28 + _b[hs_privatexaci]*(0*28) + (_b[math8] + _b[math8xaci]*28
+ _b[hs_privatexmath8]*0 + _b[hs_privatexmath8xaci]*(0*28))*x, ///
range(25 75) clcolor(orange)), ///
legend(order(1 "ACI = 7" 2 "ACI = 14" ///
3 "ACI = 21" 4 "ACI = 28")) ///
```

```
yscale(range(200 800)) xscale(range(25 75)) ytitle("Math SAT Scores") xtitle("Eighth Grade
Standardized Math Scores")

*Graph for Private Schools*
graph twoway ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*7 + _b[hs_privatexaci]*(1*7) + (_b[math8] + _b[math8xaci]*7 +
_b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*7))*x, ///
range(25 75) clcolor(blue)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*14 + _b[hs_privatexaci]*(1*14) + (_b[math8] + _b[math8xaci]*14
+ _b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*14))*x, ///
range(25 75) clcolor(red)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*21 + _b[hs_privatexaci]*(1*21) + (_b[math8] + _b[math8xaci]*21
+ _b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*21))*x, ///
range(25 75) clcolor(green)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*28 + _b[hs_privatexaci]*(1*28) + (_b[math8] + _b[math8xaci]*28
+ _b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*28))*x, ///
range(25 75) clcolor(orange)), ///
legend(order(1 "ACI = 7" 2 "ACI = 14" ///
3 "ACI = 21" 4 "ACI = 28")) ///
yscale(range(200 800)) xscale(range(25 75)) ytitle("Math SAT Scores") xtitle("Eighth Grade
Standardized Math Scores")

*Graph for Private Schools Looking At Intersection*
graph twoway ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*7 + _b[hs_privatexaci]*(1*7) + (_b[math8] + _b[math8xaci]*7 +
_b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*7))*x, ///
range(55 90) clcolor(blue)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*14 + _b[hs_privatexaci]*(1*14) + (_b[math8] + _b[math8xaci]*14
+ _b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*14))*x, ///
range(55 90) clcolor(red)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*21 + _b[hs_privatexaci]*(1*21) + (_b[math8] + _b[math8xaci]*21
+ _b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*21))*x, ///
range(55 90) clcolor(green)) ///
(function y = _b[_cons] + _b[ses]*3.32 + _b[female]*0.53 + _b[nwhite]*0.27 + _b[pared]*0.49
+ _b[hs_private]*0 + _b[aci]*28 + _b[hs_privatexaci]*(1*28) + (_b[math8] + _b[math8xaci]*28
+ _b[hs_privatexmath8]*1 + _b[hs_privatexmath8xaci]*(1*28))*x, ///
range(55 90) clcolor(orange)), ///
legend(order(1 "ACI = 7" 2 "ACI = 14" ///
3 "ACI = 21" 4 "ACI = 28")) ///
```

```
yscale(range(500 1300)) xscale(range(55 90)) ytitle("Math SAT Scores") xtitle("Eighth Grade
Standardized Math Scores")

*Determines if the effect of math8 and aci is statistically significant for private schools
lincom math8xaci + hs_privatexmath8xaci
*Determines if the effect of math8 and aci is statistically significant for public schools
lincom math8xaci + hs_privatexmath8xaci*0
*Determines is the effect of math8 is statistically significant for public schools
lincom math8 + hs_privatexmath8xaci*0
*Determines is the effect of math8 is statistically significant for private schools
lincom math8 + hs_privatexmath8xaci

*****
*Sensitivity Analysis
*Creates predicted math SAT scores
predict satmhat, xb
*Creates math SAT scores studentized residuals
predict satmstud, rstud
*Creates Cook's D's for math SAT scores
predict satmcd, cooksd
*Creates squared residuals of SAT math scores
gen satmstudsq= satmstud^2
*Produces scatter plot fo math SAT scores studentized residuals vs. eighth grade math scores
twoway (scatter satmstud math8) (function y=0, range(20 80))
*Produces scatter plot of math SAT scores studentized residuals vs. eighth grade math scores
twoway (scatter satmstud math8) (function y=0, range(20 80)), by (hs_private)
*Produces scatter plot of math SAT scores studentized residuals vs. aci
twoway (scatter satmstud aci) (function y=0, range(1 31))
*Produces box plot of math SAT scores studentized residuals over aci
graph box satmstud, over(aci)
*Produces scatter plot of math SAT scores residuals vs. school type
twoway (scatter satmstud hs_private) (function y=0, range(0 1))
*Produces box plot of math SAT scores over school type
graph box satmstud, over(hs_private)
*Produces scatter plot of math SAT scores vs. predicted math SAT scores
twoway (scatter satmstud satmhat) (function y=0, range (200 800))
*Produces scatter plot of math SAT scores squared studentized residuals vs. predicted math SAT
scores
scatter satmstudsq satmhat
*Produces histogram of math SAT scores studentized residuals with normal curve overlaid
histogram satmstud, normal
*Produces quantiles of math SAT scores studentized residuals against quantiles of normal
distribution
qnorm satmstud
*Produces a scatter plot of math SAT scores Cook's D's vs. math SAT scores studentized
residuals
```

```
twoway (scatter satmcd satmhat, mlabel(studentid)) (function y=4/600, range(200 800))
*Sorts math SAT scores Cook's D's values numerically from largest to smallest in the data editor
gsort -satmcd
*Lists all students with high Cook's D
list studentid satm math8 aci ses hs_private pared nwhite satmstud satmcd in 1/10
*runs a regression of SAT math scores on eighth grade math stadardized test scores, high school
academic curriculum intensity, status of high school
*the product of standardized 8th grade math scores and high school academic curriculum
intensity, the product of high school type and eighth grade math scores
*the product of high school type and high school academic curriculum intensity, and the
interaction term of high school type, eighth grade math scores, and academic
*curriculum intensity
regress satm math8 ses female nwhite pared aci hs_private math8xaci hs_privatexmath8
hs_privatexaci hs_privatexmath8xaci
**runs a regression of SAT math scores on eighth grade math stadardized test scores, high
school academic curriculum intensity, status of high school
*the product of standardized 8th grade math scores and high school academic curriculum
intensity, the product of high school type and eighth grade math scores
*the product of high school type and high school academic curriculum intensity, and the
interaction term of high school type, eighth grade math scores, and academic
*curriculum intensity without studentid=1495450
regress satm math8 ses female nwhite pared aci hs_private math8xaci hs_privatexmath8
hs_privatexaci hs_privatexmath8xaci if studentid!=1495450

*****
*drops all open graphs
graph drop _all
*clears data set
clear
*saves results
*****
```