College Admissions as Seen Through MCAS Scores

By:

Ariel Khait

1007854047

ariel.khait@mail.utoronto.ca

ECO375: Applied Econometrics

University of Toronto

Department of Economics

November 29th 2023

Abstract:

Standardized Tests and more specifically, the SAT's have major implications for the college admission decision. Relying on Freedman's 2003 study, we look to study the relationship between a different standardized test (MCAS) and college admissions. Our investigation reveals a notable association between 10th-grade MCAS scores and the enrollment count of students in colleges or universities per school. To broaden this analysis, we delve deeper into the examination of admissions, focusing on both private and public four-year institutions. The analysis examines educational covariates and demographic factors within schools, revealing disparities in the components that drive admissions to private versus public institutions.

1. Introduction

Since its inception in 1926, the SATs have served as a pivotal benchmark for gauging college readiness. Colleges commonly utilize this standardized test, in tandem with GPAs, to dictate the admission decisions of millions of high school students annually. Freedman's 2003 study established a correlation ranging from 0.725 to 0.813 between SAT and MCAS scores, significant at the 1 percent level. Following upon Freedman's 2003 study, our study aims to single out the potential influence of standardized test scores, more specifically the MCAS, on admissions into institutions of higher learning.

The Massachusetts Comprehensive Assessment System (MCAS) is a standardized test administered to Massachusetts students from 2nd grade to 10th grade. In Massachusetts, passing the 10th grade MCAS test is required to earn a high school diploma. We anticipate that higher MCAS scores would correlate with increased admissions to higher learning institutions. Through our analysis, we uncovered a significant association between MCAS scores and admission to colleges or universities. Furthermore, we extend our findings indicating disparities in the impact of MCAS scores on admissions to private versus public institutions.

2. The Context and Data

Our data for this analysis is sourced from the Massachusetts Department of Education, specifically from the 2016-17 school year. It follows a cross-sectional data structure, comprising over 1,800 observations. These observations pertain to schools in Massachusetts. However, certain schools, such as elementary and middle schools, do not encompass our unit of observation, which is 10th-grade MCAS scores. Consequently, we filtered the data to exclude non-high schools or schools lacking 10th-grade data, resulting in 402 observations. However, seven of the remaining entries lacked proper data and were consequently removed, reducing the final count to 396.

The MCAS test comprises Math and English sections, and we merge these to derive an overall average score. Additionally, we transformed the MCAS scores to a natural log scale. This conversion allows us to observe percentage changes in scores, considering the MCAS's scaling nature. For our analysis, we focus on 10th-grade standardized test scores per school. We evaluate the mean 10th-grade MCAS test scores per school, averaging 334.36, with a standard deviation of 234.21 as seen in Table 1 and Graph 1. The significant standard deviation in MCAS scores indicates a broad range of scores, contributing significantly to the precision and accuracy of our analysis. Furthermore, we examine the count of students from each school enrolled in colleges or universities, as indicated in Table one, averaging 143 students attending college, with a standard deviation of 100.46 students. Expanding on those statistics, we observe that the average percentages of students attending private four-year schools and public four-year schools are approximately 32.55% and 35.238% respectively. Although these percentages may appear modest initially, it's crucial to acknowledge the existence of numerous alternatives to a traditional four-year college education, including trade schools, two-year colleges, and individuals choosing not to pursue a college degree. Moreover, the standard deviations for these percentages are 14.3 and 13.2 respectively, indicating substantial variability among schools.

Looking at our confounding factors we have categorized them into two categories: educational covariates, and race & gender covariates. The educational covariates are comprised of variables that measure high-school specific decisions on educational outcomes. This includes metrics such as the average dollar amount spent per student for each school (E/S), student teacher ratio per school (STR), and average teacher's salary per school (ATS). This section also includes covariates that describe the student population aside from their race. This includes the number of English language learners (ELL) per school, the number of high needs (HighN) students per school, the number of disabled students (Dis) per school, and the number of students that are deemed "economically disadvantaged" (EcoDis) by the Massachusetts Department of Education. Regarding monetary variables like average expenditure per student and average salary, we examined the natural logarithm of these variables. This transformation is employed because a one-dollar variance in salary will not significantly affect college admissions, whereas a one-percent difference in salary could yield more substantial results. The race and gender covariates are all depicted as percentages, indicating the respective proportions within the school. These include African American, Native American, Native Hawaiian, Hispanic, White, Male, and multiracial student proportions, representing the diversity within the school population. We intentionally excluded the percentage of female students from our analysis due to perfect multicollinearity, as it is inherently represented by the complement of the percentage of male students within the school.

3.1 Simple Linear Regression

Initially, we examine the foundational model which is the linear relationship between the natural log of the average 10th grade MCAS scores per school and the natural log of the number of students going to college or university for each school. We estimate the relationship between a one percent change in MCAS scores and the percentage of students going to college or university as specified by model (1) or the following:

$$\ln(College\ Bound\ Students) = \beta_0 + \beta_1 \ln(10th\ Grade\ MCAS\ Scores) + u$$

To start our analysis, we would like to point your attention towards how well the model fits the data. The initial R squared value of our foundational model is 0.77 indicating that 77 percent of the variance in the natural logarithm of college admission is accounted for by the natural logarithm of 10th-grade MCAS scores. The large value of R squared already suggests that MCAS scores might serve as an effective estimator of college admissions. Furthermore, Beta one is 0.84 meaning a one percentage point increase in the average 10th grade MCAS score per school results in a 0.84 percentage point increase in the number of students going to college or university on average per school. This allows us to reject the null hypothesis that a one percent change in average 10th grade MCAS scores has no impact on college admissions ($H_0\colon \beta_1 = 0$) at a 1 percent significant level. Unfortunately, because this is a basic linear model with no confounding factors, it is subject to omitted variable bias which violates the first least squared assumptions. Looking at Table 1, we also notice that while outliers are rare, there are definitely a few in the data. The OLS estimator is very sensitive to outliers which can impact our estimation, failing another least squares assumption.

We continue our analysis with the clustering of standard errors which aims to mitigate the impact of shared characteristics within each school district. Continuing from our base model we cluster standard errors by school district codes as seen by Table 2 Model (2). Districts often exhibit distinct characteristics that could influence 10th-grade MCAS scores. An example of this would be a school district mandating that a school be 6 and a half hours instead of the state mandated 6 hours. If two schools were within the same school district, both schools would be expected to have a higher average 10th grade MCAS score due to the increased schooling in those districts. By clustering at the school district level, we look at the aggregate of these two schools which allows for a more accurate estimation of the impact of average 10th grade MCAS scores. Adding clustered standard errors, as seen in Table 2, doesn't impact the R squared or the value of the regression but does decrease the value of the standard error from 0.0651 to 0.0639 which increases the accuracy of our regression and minimizes the 95 percent confidence interval.

Shifting our focus to the relationship between average 10th-grade MCAS scores and the percentage of students admitted into private versus public 4-year colleges respectively. Regression models (1) and (2) in Table 3, estimate the percentage point change in students attending a public or private 4-year school induced by a one percent change in MCAS scores. The models are specified as such:

$$\ln(Private\ College\ Bound\ Students) = \beta_0 + \beta_1 \ln(10th\ Grade\ MCAS\ Scores) + u$$

$$\ln(Public\ College\ Bound\ Students) = \beta_0 + \beta_1 \ln(10th\ Grade\ MCAS\ Scores) + u$$

Studying the fit of these two regressions, we can see that the base models don't describe the data adequately. Only 3.3 percent of the variation in the percentage of private 4-year college admissions is fit by average 10th grade MCAS scores while 6.6 percent of the variation in the percentage of admissions to public 4-year schools is fit by average 10th grade MCAS scores. We used clustered standard errors to eliminate any covariation within the districts. Looking at the two Beta one's of their regressions we can see that a one percent increase in the average 10th grade MCAS scores had a 2.95 percent increase in students admitted to 4-year private colleges and a 3.7 percent increase in students admitted to 4-year public schools. Both results are significant at the 1 percent level, meaning they reject the null hypothesis that a one percent increase of the average 10th grade MCAS scores doesn't influence 4 years of private or public-school admissions. Since these models are a more specified version of the model in Table 2, these basic models suffer from the same drawbacks of omitted variable bias and outliers. Specifically, looking at Graph 2, there are 7 schools that did not report the percentage of students going to a public 4-year university as seen by the blue points on the X-axis. This is a major failure of the third least squares assumption as the OLS estimator is heavily influenced by outliers.

3.2 Multiple Regression

We continue our analysis by adding confounding race factors, eliminating the bias caused by the demographics of different schools. The model is specified by:

$$\%\ of\ Admission\ to\ Private\ (model\ 3)\ or\ Public\ (model\ 4)\ College = \beta_0 + \beta_1 \ln(10th\ grade\ MCAS) + \beta_2(\%\ African\ American) + \beta_3(\%\ Hispanic) + \beta_4(\%\ Native\ American) + \beta_5(\%\ Multi - Race) + \beta_6(\%\ White) + \beta_7(\%\ Male) + u$$

Focusing on model (3) in Table (2), we can see that there is a 0.03 percentage point difference in the regression value and a decrease in standard error which help improve the accuracy of the model but not by a significant margin. However, the R squared of the model jumps from the 0.77 specified by model (2) to 0.83 meaning 6 percent more of the variation in the natural log of college admissions is fit by the race and gender covariates in model (3). Extending this analysis, we can add educational covariates to further improve our R squared. In model (4), our regression specifies that a 1 percent change in the average 10th grader MCAS score leads to a 0.76 percent change in the average number of students going to college or university per school. Unfortunately, the standard errors almost doubled from 0.0633 to 0.132 but our beta value is still statistically significant at 1 one percent level when testing for no effect ($\beta_1 = 0$). Also, it is important to note that once we added the educational covariates, the R squared increased from 0.83 to 0.86 which is not a large increase but does improve the fit of the variation of the natural log of average students going to college per school. Overall, the addition of education covariates explains more of the data and helps eliminate some positive bias but at the cost of the accuracy of our estimate, more specifically doubling our standard errors.

Further dissecting college admissions into private versus public schools provides intriguing insights into the varying impact of MCAS scores on these distinct educational institutions. Referring to Table 3 presents the regressions comparing these institutions with clustered standard errors. Mirroring our previous methodology, we first add the race covaries which decrease our regression value from 2.95 to 1.95 in model (3) and from 3.7 to 2.48 in model (4). This suggests that models (1) and (2) were exhibiting positive bias due to race and gender. The standard errors in both cases stay within a 0.025 range of their original value and are therefore an insignificant change. With that being said, the large decrease in the regression value of model (3) leads to it being significant at the 5 percent level as opposed to the 1 percent level from before. In model (4) while the introduction of race and gender to the model does decrease the value of the regression, it is still significant at the 1 percent level. In context, a percentage point increase in average 10th grade MCAS scores per school increases the percentage of students going to private schools by 1.95 percent and the percentage of students going to public schools by 2.48 percent once race and gender are introduced as confounding factors in the model. Furthermore, the R-squared of models (3) and (4) jump up to 0.30 and 0.37 respectively, as compared to the R squares of the simple models which were 0.03 and 0.07. This means that the race and gender factors fit 27 percent of the variation in the percentage of students going to a 4-year private college (model (3)) and 30 percent of the variation in the percentage of students going to a 4-year public college (model (4)).

Concentrating on the percentage of males in different schools, we notice a significant effect on admissions to private and public schools. As seen in models (3) and (4), we notice a 1 percent increase in the percentage of 10th grade males in Massachusetts is associated with a 0.8 percent decrease in the percentage of students going to a private or public 4-year school which is significant at the 1 percent level. This implies that the number of males does have an impact on the number of students that go to college from that school. In conclusion, the increase in R squared coupled with the notable influence of the male variable on the percentage of students attending public versus private schools leads us to deduce that models (3) and (4) offer a more comprehensive explanation for the correlation between average 10th-grade MCAS scores and college admissions to private versus public schools. However, these relationships are not as large as observed in the previous models due to some positive bias being rooted out as a result of these confounding factors.

Extending our previous models, we add different educational factors as defined previously. We then specify our final model as the following:

$$\%\ of\ Admission\ to\ Private\ (model\ 5)\ or\ Public\ (model\ 6)\ College = \beta_0 + \beta_1 \ln(10th\ grade\ MCAS) + \beta_2(\%\ African\ American) + \beta_3(\%\ Hispanic) + \beta_4(\%\ Native\ American) + \beta_5(\%\ Multi-Race) + \beta_6(\%\ White) + \beta_7(\%\ Male) + \beta_8 ln(E/S) + \beta_9 ln(ATS) + \beta_{10}(Average\ Class\ Size) + \beta_{11}(ELL) + \beta_{12}(Dis) + \beta_{13}(HighN) + \beta_{14}(EcoDis) + u$$

With these new additions we see the regression jump to 3.4 in model (5) and 3.23 in model (6) suggesting that we were exhibiting negative bias in models (3) and (4) due to education factors. We can also infer that the negative bias caused by the education factors was similar in magnitude to the positive bias exhibited by race and gender covariates due to the regressions being similar to models (1) and (2). We start our analysis by looking at the R-squared values of models (5) and (6) which are 0.535 and 0.518 respectively. This implies that 24 percent more of the variance in the percentage of students going to a 4-year private school can be fit by model (5) as compared to model (3). Similarly, 15.2 percent more of the variation in the percentage of students attending a 4-year public school can be fit by model (6) as compared to model (4). Factoring in the actual effect of the education factors on our model, we see the first major difference between the two models.

Focusing on the percentage of students admitted to private schools, most of the impact on admission is caused by the percentage of males per high school, average teacher's salary, and wealth (economically disadvantaged metric). Observing Model (5), It is also important to note that the number of high-needs students per school does have a significant effect at the 10 percent level, but it is a small effect of 0.037 which doesn't impact our model significantly. Upon closer examination of the educational factors, it becomes evident that both the natural logarithm of the average teacher's salary and the percentage of economically disadvantaged students are important factors as they are both statistically significant at the 1 percent level. Drawing attention to the mean salary per teacher, a one-percentage-point rise in the average teacher's salary corresponds to a 28.41-percentage-point increase in the percentage of students admitted to a 4-year private school. This outcome bears substantial significance as a 28-percentage point increment represents twice the standard deviation of the percentage of students enrolled in a 4-year private school. This might be attributed to the limited number of observations, specifically 297, within this model. Continuing with model (5), we can see that a one-student increase in economically disadvantaged students per school negatively impacts the percentage of students going to a 4-year private school by 0.067 percent at the one percent significance level. We can conclude from model (5) that since private schools are on average $27,847 more expensive than public schools (US News) we see that wealth-impacted factors such as the economically disadvantaged scale, and average teachers' salaries in school districts significantly impact admissions to a 4-year private college.

Shifting our focus to admissions in public 4-year colleges, we observe that school-wide covariates like class size, expenditure per person, and the presence of English language learners exert a more substantial impact on admissions compared to factors associated with wealth, as observed in previous models. Analyzing Model (6), we note that despite retaining significance at the 1 percent level, a one percent rise in the count of males per school results in a 0.6 percentage point decline in the number of students admitted to a 4-year public school, marking a 0.2 percentage point increase over the effect depicted in model (4). Furthermore, we see that a one percent increase in expenditure per person decreases the percentage of people attending a 4-year public school by 10.34 percent which is significant at the 1 percent level. Moving downwards we see that the average class size and the number of English language learners are both significant at the 1 percent level with coefficients of 0.87 and 0.04 respectively. A one-student increase in the average class size corresponds with a 0.87 percentage point increase in students going to a public 4-year college. Additionally, a one-person increase in the number of English language learners in a school lead to a 0.04 percentage point increase in the percentage of students attending a 4-year public school.

4. Limitations of results

Looking at our final models in Table 3, we can see the impact of the efficiency-bias trade-off where the standard errors of our regressions which were previously below 1 are now around 1.5 causing our values to be significant at the 5 percent level instead of the 1 percent level. This suggests that we don’t have enough data to lower standard errors to a point where we can root out any bias while confidently stating that average 10th-grade MCAS scores have a substantial effect on college admissions.

Additionally, we’d like to emphasize that the primary goal of the MCAS is to inform educators and parents about their children's academic standing. A lower score on this assessment might prompt increased consideration for private schooling, particularly in affluent areas. This trend is notably prevalent because the MCAS is administered almost a year before the SATs. This timing provides parents whose children are classified as "less than proficient" by the MCAS an opportunity to bridge gaps and potentially alter their educational trajectory. This scenario highlights the presence of an omitted variable bias - the access to tutoring.

When evaluating the validity of our data, we encountered some evident threats to both internal and external validity. We did not have any direct statistics about family income besides the economically disadvantaged metric, which would have had a substantial effect on admissions into private schools. Private schools, as seen by US News, are substantially more expensive than public schools which threaten internal validity.

Regarding external validity, this study's findings can be extended to various other states like Texas, California, and Florida, all of which conduct their own form of statewide standardized tests. Despite potential demographic variations, our model demonstrates that demographic differences don't notably affect the model. However, it's crucial to acknowledge the diverse wealth distribution among states. For instance, California might have a different wealth distribution compared to Massachusetts, so models heavily influenced by wealth may yield disparate results in California.

5. Conclusion

Our study explored the relationship between standardized test scores, specifically the Massachusetts Comprehensive Assessment System (MCAS), and college admissions. Analyzing data from Massachusetts schools, we discovered a significant correlation between higher MCAS scores and increased admissions to both private and public four-year institutions.

While shedding light on the impact of MCAS scores on college admissions, our study also highlighted limitations, such as fluctuations in standard error and the absence of direct family income data. Nevertheless, our findings reinforce the relevance of standardized test scores as crucial indicators in the college admission process.

In essence, our research contributes to understanding the influential role of MCAS scores in predicting college admission readiness, emphasizing their importance as assessment metrics for college admissions.

References:

Freedman, Nathaniel, and John Wilkes. OPPOSITE OBJECTIVES, SIMILAR SCORES: THE MCAS and SAT an Interactive Qualifying Project Report. 2007.

Kerr, Emma, and Sarah Wood. "The Cost of Private vs. Public Colleges." US News & World Report, U.S. News & World Report, 8 June 2022, www.usnews.com/education/best-colleges/paying-for-college/articles/2019-06-25/the-cost-of-private-vs-public-colleges.

Lofaro, Angela. The Growth of the Test-Optional Movement: Analysis of Test- the Growth of the Test-Optional Movement: Analysis of Test- Optional Admissions Policies in American Higher Education Optional Admissions Policies in American Higher Education. 2021.

Oehrlein, Paul. Digital Commons @ IWU Digital Commons @ IWU Determining the Future Income of College Students Determining the Future Income of College Students. 2009.

Table 1:

Variable	Obs	Mean	Std. Dev.	Min	Max
MCAS Combined	350	334.36	234.2103	3	1518
Students to College/Univ	341	143.088	100.4619	2	682
% of Students to Private Schools	341	32.55044	14.31634	0	77.8
% of Students to Public Schools	341	35.23871	13.17497	0	67.9
Educational Covariates					
Average Expenditure Per Student	351	15414.1	2929.918	10399.97	28207.95
Average Teachers Salary	351	74339.63	8330.683	34588	100731
Average Class Size	384	15.40156	4.073765	3.6	34
English Language Learner	396	52.76515	108.914	0	1020
Students With Disabilities	396	130.4672	93.50238	0	589
High Needs Students	396	329.5051	323.9671	10	2646
Economically Disadvantaged	396	222.4394	253.8916	4	1997
Race and Gender Covariates					
African American	396	11.59268	16.9813	0	80.5
Native Hawaii	396	.1017677	.2507142	0	3.4
Hispanic	396	19.06742	21.59354	0	95.5
Native American	396	.2333333	.4195839	0	5.1
Multi – Racial	396	2.74697	1.896063	0	13.2
White	396	61.76742	31.6375	0.6	97.9
Asian	396	4.491162	6.609039	0	58.4
Male	396	51.98586	7.553955	24	88.6

Table 2:

	(1) Attending College	(2) Attending College	(3) Attending College	(4) Attending College
Ln(10^{th} Grade MCAS Scores)	0.840***	0.840***	0.812***	0.764***
	(0.0651)	(0.0639)	(0.0633)	(0.132)
Clustered SE	No	Yes	Yes	Yes
Race FE	No	No	Yes	Yes
Educational FE	No	No	No	Yes
R-Squared	0.770	0.770	0.828	0.863
Number of Schools	331	331	331	297

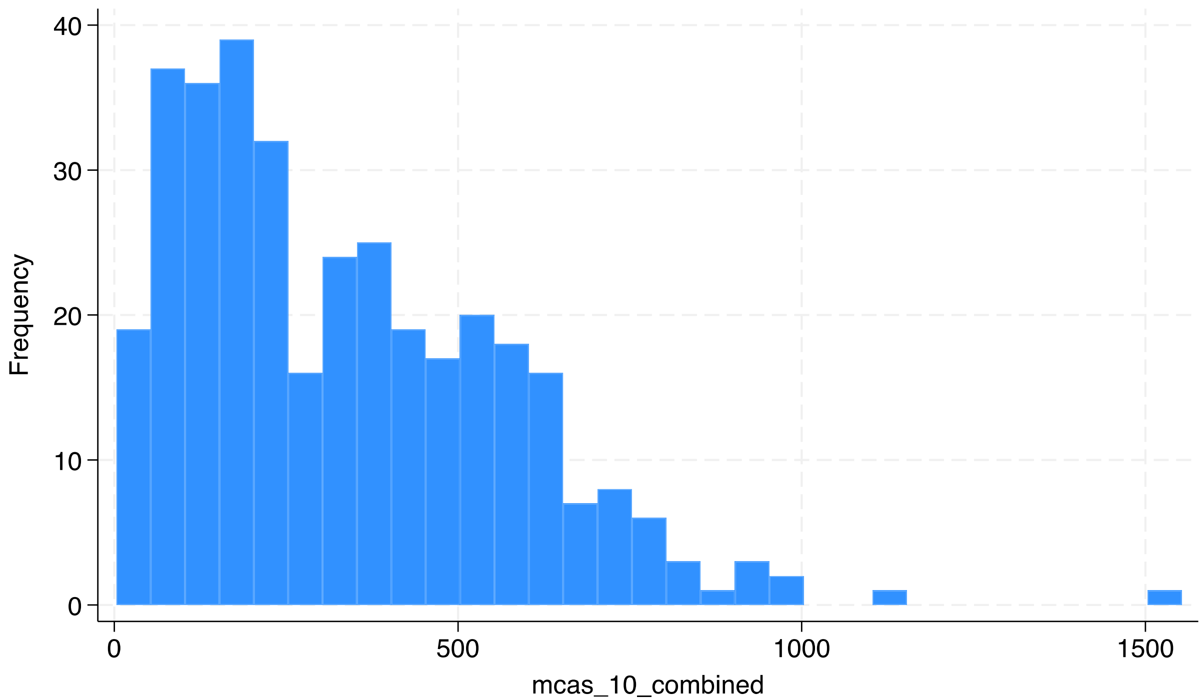
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Table 3:

	(1) % Students to 4yr Private	(2) % Students to 4yr Public	(3) % Students to 4yr Private	(4) % Students to 4yr Public	(5) % Students to 4yr Private	(6) % Students to 4yr Public
Ln(10th Grade MCAS Scores)	2.954***	3.695***	1.945**	2.476***	3.398**	3.236**
	(0.949)	(0.885)	(0.968)	(0.907)	(1.575)	(1.519)
% African American			1.061	3.178	-1.812	-2.689
			(8.127)	(6.699)	(6.809)	(6.516)
% Native Hawaiian			5.084	7.105	3.895	-0.625
			(8.631)	(8.430)	(8.132)	(7.507)
% Hispanic			0.736	2.916	-1.874	-2.930
			(8.121)	(6.711)	(6.796)	(6.534)
% Native American			0.587	5.277	-3.077	-0.309
			(8.004)	(6.544)	(6.700)	(6.379)
% Multi-Race			1.194	1.497	-1.101	-3.329
			(8.206)	(6.704)	(6.874)	(6.534)
% White			0.990	3.168	-1.785	-2.663
			(8.117)	(6.709)	(6.809)	(6.529)
% Asian			1.422	3.309	-1.507	-2.615
			(8.089)	(6.725)	(6.812)	(6.545)
% Males			-0.803***	-0.823***	-0.860***	-0.605***
			(0.141)	(0.142)	(0.203)	(0.146)
Ln(Expenditure per Person)					4.500	-10.34***
					(5.010)	(3.651)
Ln(Average Salary)					28.41***	6.579
					(7.479)	(6.594)
Average Class Size					-0.201	0.866***
					(0.339)	(0.223)
English Language Learner					0.0131	0.0437***
					(0.0179)	(0.0105)
Students With Disabilities					-0.0136	-0.00334
					(0.0187)	(0.0101)
High Needs					0.0373*	-0.0152
					(0.0224)	(0.0123)
Economically Disadvantaged					-0.0671***	-0.00117
					(0.0220)	(0.0117)
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.0334	0.0662	0.295	0.366	0.535	0.518
Number of Schools	331	331	331	331	297	297

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Graph 1:



Graph 2: