The Effects of Class Size on Academic Success

Eric Pan

November 2022

1 Introduction

Academic success is one of the most important factors for ensuring that young people grow up to become economically successful and engaged members of society. A large part of this academic success depends on the quality of students' schooling, especially in secondary education, where performance in classes and on exams such as the SAT or ACT may have a large impact on whether or not they go college as well as the quality of the college they attend. Present literature supports the existence of a positive impact of small class size on academic performance (Arias & Walker, 2010; Nye et al., 2000). Moreover, for schools with limited resources, it is important to know whether reducing class size would serve more benefit for students as compared to other policy choices.

2 Methods

To determine the effect of class size on academic success, I will use the data contained in the *star.dta* dataset, which contains observations recording BLAH. In my analysis, I will control for factors including teacher experience and teacher's degree of education. I will run a linear regression to determine the effect of class size on math, reading, and listening scores on the SAT, and to determine whether this has a more meaningful impact on students' academic success when compared to teacher quality. Additionally, we will see on which subjects these factors play the largest role when it comes to student success.

In my analysis, I have controlled for teacher experience as well as teachers' education. With regards to the former, I have created a new variable indicating all teachers within one standard deviation of teaching experience, which is measured in years teaching. My analysis is focused on students in classrooms with teachers having an average amount of experience and without master's degrees.

3 Results

Let us begin with the following regression table depicting the regression of math on class size:

Source	SS	df	MS	Number of obs	=	2,153
				F(1, 2151)	=	24.46
Model	56078.8514	1	56078.8514	Prob > F	=	0.0000
Residual	4930986.23	2,151	2292.41573	R-squared	=	0.0112
				Adj R-squared	=	0.0108
Total	4987065.08	2,152	2317.40943	Root MSE	=	47.879

math	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
class_size	-1.26374	.2555082	-4.95	0.000	-1.764809	-.762671
_cons	511.773	5.355648	95.56	0.000	501.2702	522.2758

First, we can see that our results are significant at the $p < 0.001$ level. The coefficient tells us that on average, an extra student in the classroom correlates with an approximately 1.25 point decrease in math SAT score. However, because R-squared is only 0.0112, not much of the variance in these scores is caused by the classroom size. In fact, the average score for math was approximately 486, and so the benefit of smaller class size on math scores is fairly negligible.

Now, let's look at a similar regression table, this time for reading scores:

Source	SS	df	MS	Number of obs	=	2,153
				F(1, 2151)	=	26.17
Model	26846.9999	1	26846.9999	Prob > F	=	0.0000
Residual	2206388.23	2,151	1025.74999	R-squared	=	0.0120
				Adj R-squared	=	0.0116
Total	2233235.23	2,152	1037.74871	Root MSE	=	32.027

read	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
class_size	-.8743916	.1709145	-5.12	0.000	-1.209566	-.5392168
_cons	454.9172	3.582499	126.98	0.000	447.8917	461.9427

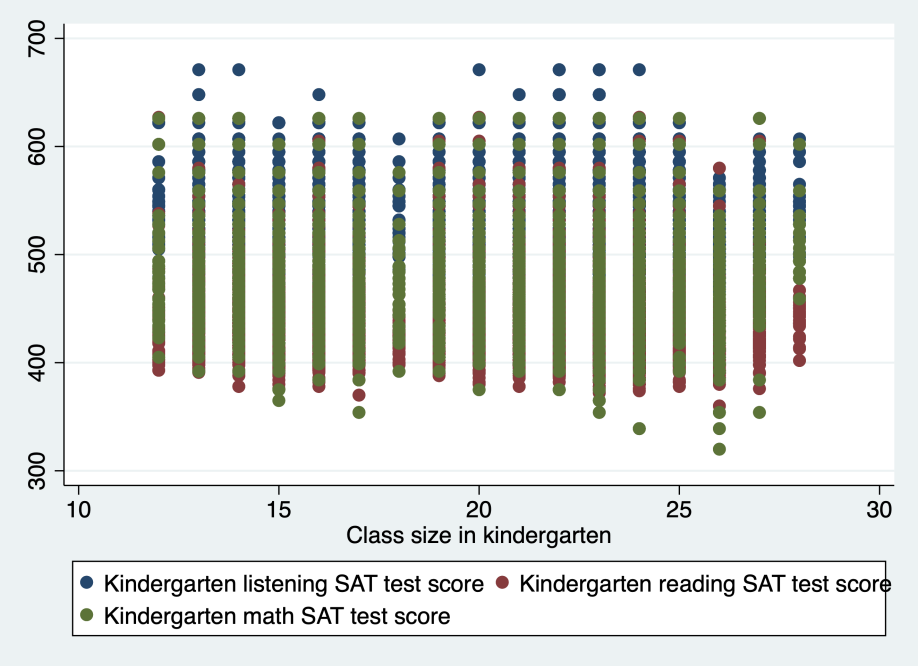
These results look very familiar — the regression coefficient gives an approximately 0.9 point decrease per extra student, but we have a comparable R-squared as in our math regression, and the average reading score was 437, suggesting a similarly negligible effect of class size. Finally, let's take a look at the regression table for listening:

Source	SS	df	MS	Number of obs	=	2,153
				F(1, 2151)	=	17.99
Model	18468.3454	1	18468.3454	Prob > F	=	0.0000
Residual	2208128.38	2,151	1026.55899	R-squared	=	0.0083
				Adj R-squared	=	0.0078
Total	2226596.72	2,152	1034.6639	Root MSE	=	32.04

listen	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
class_size	-.7252239	.1709819	-4.24	0.000	-1.060531	-.3899169
_cons	550.9555	3.583912	153.73	0.000	543.9272	557.9837

Again, it does look like a significant negative effect of class size on score, though the effect is rather small. It is notable that the effect of class size on math score was larger than that on reading and listening, however, lending some credence to the effectiveness of smaller classes, especially for more technical subjects such as math.

Now let's compare these three subjects on a scatter plot:



As expected, no clear trends appear when comparing the three, though it does seem that overall students do better on listening than on other subjects.

Finally, let's take another look at the math regression, but with teachers who have earned master's degrees:

Source	SS	df	MS	Number of obs	=	1,716
				F(1, 1714)	=	18.62
Model	40419.1425	1	40419.1425	Prob > F	=	0.0000
Residual	3721107.33	1,714	2171.00777	R-squared	=	0.0107
				Adj R-squared	=	0.0102
Total	3761526.47	1,715	2193.30989	Root MSE	=	46.594

math	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
class_size	-1.200154	.2781467	-4.31	0.000	-1.745696	-.6546108
_cons	509.7827	5.679322	89.76	0.000	498.6435	520.9218

The table here looks functionally identical to that of the other regression of math — it seems as though more qualified teachers have no particular advantage when it comes to teaching students in smaller classes.

4 Conclusion

From our results, we have determined that although class size has a significant effect on students' academic performance, the effect is relatively small, and the difference in performance is not greatly affected by class size overall. It is worth noting, however, that more well-funded schools may be able to better support smaller classes, or schools serving more students may have more funding which can be used to support students in other ways. It is worth investigating other factors contributing to student performance, as reducing the number of students per class is not likely to grant students a large advantage on exams such as the SAT.

In this exercise, I learned the complete process of formulating a research question based on given data and analyizing it in different ways to make conclusions.