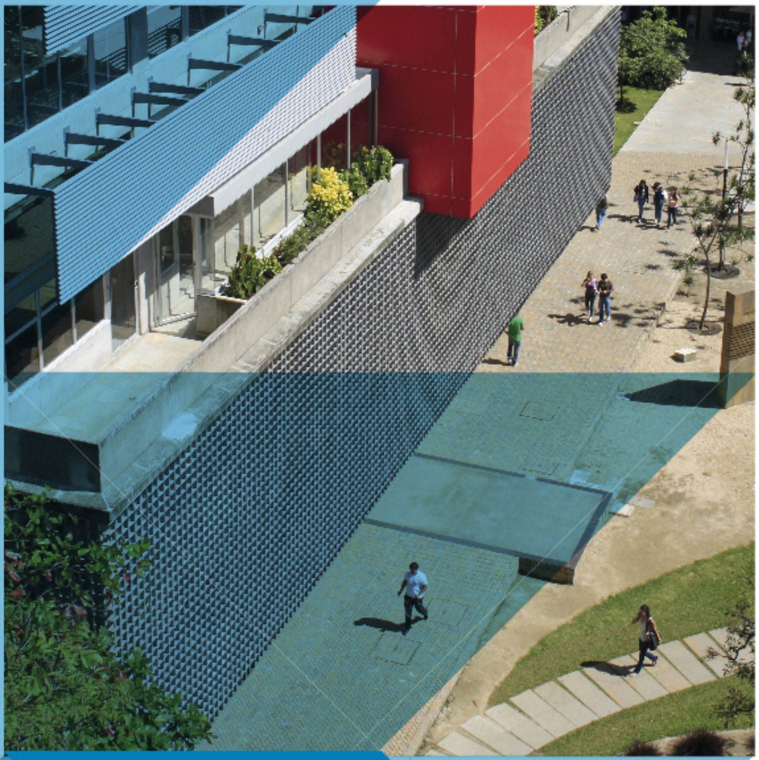


UNIVERSIDAD EAFIT
Abierta al mundo

Escuela de Economía y Finanzas

Documentos de trabajo

Economía y Finanzas

Centro de Investigación Económicas y Financieras

No. 12-05
2013

Evidence that Class Size Matters in 4th Grade Mathematics: An Analysis of TIMSS 2007 Data for Colombia

Breton, Theodore R.

cief
Centro de Investigaciones Económicas y Financieras
Universidad EAFIT

Evidence that Class Size Matters in 4th Grade Mathematics

An Analysis of TIMSS 2007 Data for Colombia

Theodore R. Breton*

Universidad EAFIT

tbreton@eafit.edu.co and ted.breton@gmail.com

January 11, 2013

Abstract

Like students in most developing countries, Colombian students in 4th grade performed poorly in the TIMSS 2007 test of mathematics skills, achieving an average score of 355 relative to an international mean of 500. After controlling for other factors and misreporting error, I find that large classes have substantial adverse effects on student achievement. Increases in class size from 20 to 53 students reduce test scores by about 80 points, or 2.4 points for each additional student in the class. Most likely this is the cumulative effect of class size in grades one to four on achievement in 4th grade.

JEL Codes: I20, H52

Key Words: International Education, Colombia, Class Size, Test Scores, TIMSS

*Thanks to Carlos Albert Velasquez for converting the TIMSS SPSS data to STATA format and to Andres Ramirez Hassan, Juan Carlos Duque, Isabel Fernandes (Subdirectora de Análisis y Divulgación del ICFES), and researchers at the TIMSS and PIRLS International Study Center for assistance related to the analysis of TIMSS data.

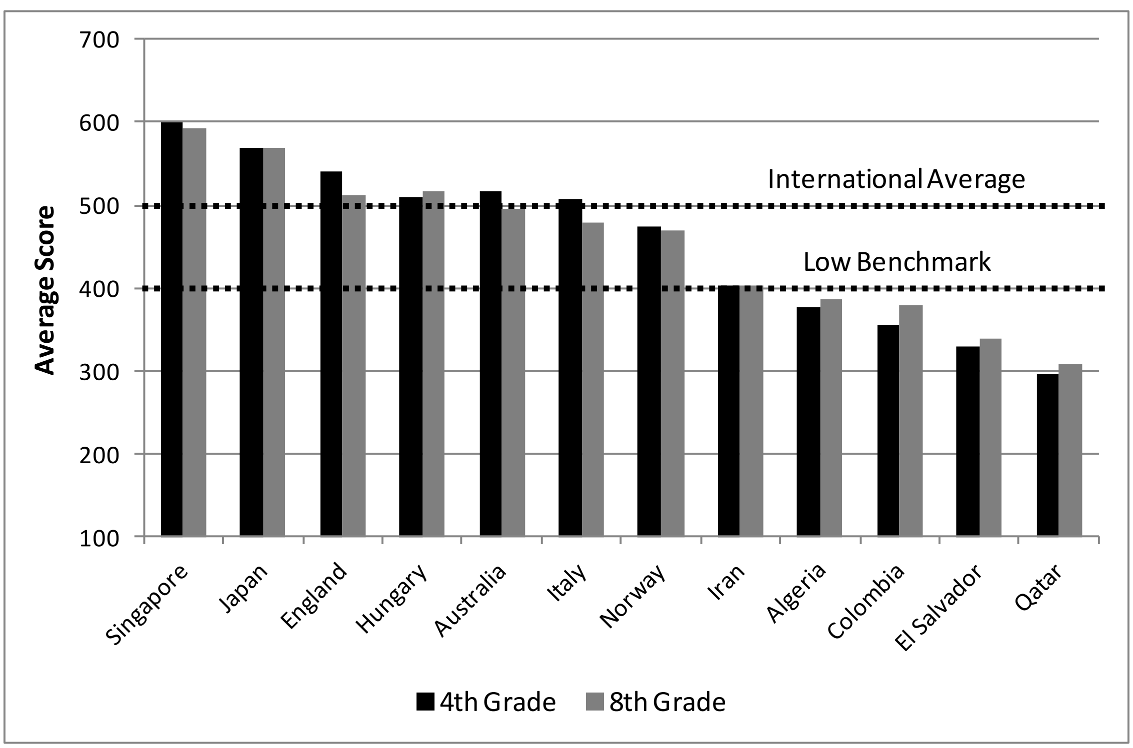
I. Introduction

There is considerable evidence that a student's cognitive skills affect his/her future personal income and that in the aggregate these skills affect future national income [Hanushek and Woessmann, 2008, and Breton, 2011]. In response to this evidence, a growing number of countries regularly test their students in national and international examinations to determine their skills. One of these examinations is the Trends in International Mathematics and Science Study (TIMSS).

Figure 1 shows the average test scores in mathematics in 4^{th} and 8^{th} grade for a subset of the countries that participated in TIMSS 2007 [Mullis, Martin, and Foy, 2008]. The average international score was 500. A score of 400 is considered a low benchmark for acceptable mathematics skills on this test [Foy and Olson, 2009].

Figure 1

Average Scores on TIMSS 2007 Tests of Mathematics Skills



Students in developing countries typically do not perform well on these tests [Hanushek and Woessmann, 2008]. As shown in the figure, the imputed TIMSS 2007 average score in mathematics for the Colombian 4th grade population is 355, which is considerably below the low benchmark of acceptable skills. In addition, the distribution of Colombian test scores indicates that 95 percent of students score below 500 [ICFES [2010].

Scores in 4th and 8th grade are highly correlated. In countries where students have low average scores in 4th grade, many drop out of school, and those remaining have low average scores in 8th grade. Since skills in mathematics are cumulative, deficiencies in 4th grade are an obstacle to later learning. The end result is that upon completion of secondary school, only a very small fraction of students are qualified to pursue careers that require a strong foundation in mathematics.

While schools and teachers are often blamed for low scores, studies across countries consistently show that family characteristics, such as the education level of the parents and family income, are an important determinant of student achievement [Parcel and Dufur, 2009]. Due to the limited educational opportunities historically for most children in developing countries, most parents have little education and little income. These limitations adversely affect their children's achievement in school.

School and teacher characteristics and teaching methods also affect student achievement, and in the near term changes in these characteristics are the only option potentially available to raise students' skills. But despite the enormous number of studies that have been carried out, there is a surprising lack of consensus on the schooling characteristics that affect student achievement. The empirical results from these studies vary between developed and developing countries, between countries within these groups, between primary and secondary schools, between studies that include different sets of characteristics, and between studies that use different statistical techniques to quantify effects.

Teachers generally believe that they can improve student achievement if the size of classes is reduced, but numerous studies have shown that the positive effect is usually small [Hattie, 2005]. Given the enormous cost of reducing class size, Hattie [2005] argues that reductions in class size are much less cost-effective than other policies as a means of raising student achievement.

Hattie [2009] summarizes over 800 meta-analyses to determine the effectiveness of numerous proposed policies to raise student achievement. He shows that the effect size of reductions in class size in these analyses is small (d= 0.1- 0.2).[1] But he cautions that these results are not generalizable to non-English-speaking or to non-highly developed countries. His summary of meta-analyses shows the effect of reductions in class size from 25 to 15 students in highly-developed countries. In contrast, the proposed reduction in class size in developing countries is typically from 30-80 students to 20-30 students. Reductions in the size of these larger classes could have a much larger effect on student achievement and could be more cost-effective than the reductions studied in developed countries.

Hattie [2009] observes that in classes of 30-80, teaching is more oriented to rote learning, while in classes of 20-30, students can be grouped and provided with differentiated instruction. Effective teaching in classes with more than 30 students requires skilled classroom management to be effective. If teachers are unable to maintain discipline in these classes, the level of student achievement can be quite low. Hattie reports that in mega-analyses the effect size of improved classroom management is very large.

A number of studies have found that large classes in primary school mathematics have a greater adverse effect on student achievement than large classes in other subjects. Rivkin, Hanushek, and Kain [2005] found adverse effects on achievement from larger classes in 4th and 5th grade in a longitudinal study in Texas, and the adverse effects were much larger in mathematics than in reading. Using TIMSS 1995 test score data, Hanushek and Luque [2003] found evidence that larger mathematics classes had a negative effect on achievement in 14 of 17 (mostly developed) countries at age 9 but not at age 13.

[1] Effect size is calculated by dividing the change in student test scores by the standard deviation of the scores.

In a very recent meta-analysis of studies performed in developing countries since 1990, Glewwe, Hanushek, Humpage, and Ravina [2011] report that most of the studies that they find acceptable (26 out of 47) and most of the studies they deem to be of highest-quality (five out of only eight) find evidence that larger classes adversely affect student achievement and that other school characteristics, such as teacher education, teacher experience, and the availability of computers generally do not. Their findings indicate that researchers should take another look at the effect of class size on student achievement in primary school in developing countries, since reductions in class size may be one of the few policy options shown to raise students' skills in these countries.

Since class sizes in Colombia tend to be much larger and scores are much lower than in developed countries, I hypothesized that the large classes are adversely affecting student achievement in 4th grade mathematics. The average class size in Colombia is considerably larger than the average class size in 16 of the 17 countries that Hanushek and Luque [2003] examined in their study using TIMSS 1995 data.

In this paper I present the results from my analysis of this effect. The results provide evidence that my *a priori* hypothesis was correct: Large classes in Colombia appear to be having a substantial adverse effect on student performance in tests of mathematics skills in 4th grade. I found that after controlling for other factors, an increase in class size from 20 to 53 students reduces students' average test scores by about 80 points, or 2.4 points for each additional student in the class. The implication is that a much higher share of students would achieve scores above 400, the low benchmark for acceptable skills, if classes were considerably smaller.

On the other hand, it seems likely that changing the class size in 4th grade alone would not be sufficient to substantially raise existing scores. Given the lack of controls for achievement prior to 4th grade and the likelihood that most students in 4th grade were previously in classes of a similar size, the estimated effects in 4th grade likely measure the *cumulative* effect of class size during grades one to four.

Another important finding from this study is that a substantial fraction (apparently 22%) of the TIMSS 2007 data on class size in Colombia were misreported. Many teachers reported the total number of 4^{th} grade students they taught in two classes as their estimate of class size. The estimated effect of class size on achievement using these data is negative but not statistically significant. In contrast, estimates of the effect of class size on achievement that exclude the misreported data, or that use the number of students taking the test as the measure of class size, are much larger and have a high level of statistical significance.

The remainder of this paper is organized as follows. Section II presents the methodology used in the analysis. Section III presents the results. Section IV concludes.

II. Methodology

A common approach for evaluating student achievement is to estimate a student´s test score (TS) as a function of the student's personal and family characteristics, the teacher's characteristics, the teaching methods employed, and the characteristics of the school environment (X_j):

(1) $TS_j = \alpha_0 + \Sigma\, \alpha_j\, X_{ij} + \varepsilon_i$

In this paper I estimate various versions of this model, with class size included as one of the characteristics of the school environment.

Each study of student achievement differs with respect to the structure of the tests, the type of data collected, and the methods used to collect the data. TIMSS 2007 provided a variety of tests to students and then estimated the score each student would have received on a uniform test. Due to the uncertainty in this estimate, TIMSS provides five plausible values for each student's test score. Details on the creation of these data are provided in Olson, Martin, and Mullis [2008].

TIMSS 2007 collected additional data using questionnaires provided to the student, the teacher, the director of the school, and a national expert on the curriculum. The questions included in these questionnaires are provided in IAEE [2007]. In this paper I include variables that utilize data obtained from the first three questionnaires.

The sample of schools participating in the TIMSS 2007 evaluation was stratified to improve the evaluation of the effects of the learning environment in less common categories of schools. The TIMSS 2007 data set includes sampling weights for use in the statistical analysis so that estimates of the effect of different characteristics on student test scores are representative of the Colombian school population. I estimated all of the models using weighted least squares (WLS) and these sample weights.

I estimated the models using STATA 10 after conversion of the data format from SPSS to STATA. I utilized a STATA program created by Macdonald [2008] to analyze the data. This program provides statistical estimates that account for the variance in the five plausible values provided for each student's imputed score and for the sample weights provided in the TIMSS data. As a test of the validity of this software, I estimated the effect of the characteristics on the average of the five plausible values using the standard STATA 10 software. As expected, the standard software produced the same estimated coefficients but smaller estimates of variance than MacDonald's program.

The TIMSS data set for 4th grade mathematics in Colombia includes observations for 4801 students. These students are divided into two types of classes, mathematics classes and integrated mathematics and science classes. I decided to use only the data for the mathematics classes to avoid bias from potential variation in the amount of time allocated to mathematics in the integrated classes. The data set for the mathematics classes included 2361 students in 66 classes. The variables used in the regressions, the unweighted mean for each variable, its range, and the number of observations for each variable are shown in Table 1.

2.1 Class Size Data

The TIMSS 2007 data include the teacher's estimate of the size of each 4th grade class. As shown in Table 1, these estimates range from 6 to 80 students, while the number of students tested ranges from 3 to 50. Since it seemed unlikely that classes had as many as 80 students and that such a large number could have been absent on the day of the test, I compared the number of students actually tested in each class to the reported size of the class. This comparison is shown in Figure 2. It reveals that many classes have a reported size that is about double the number of students who took the test.

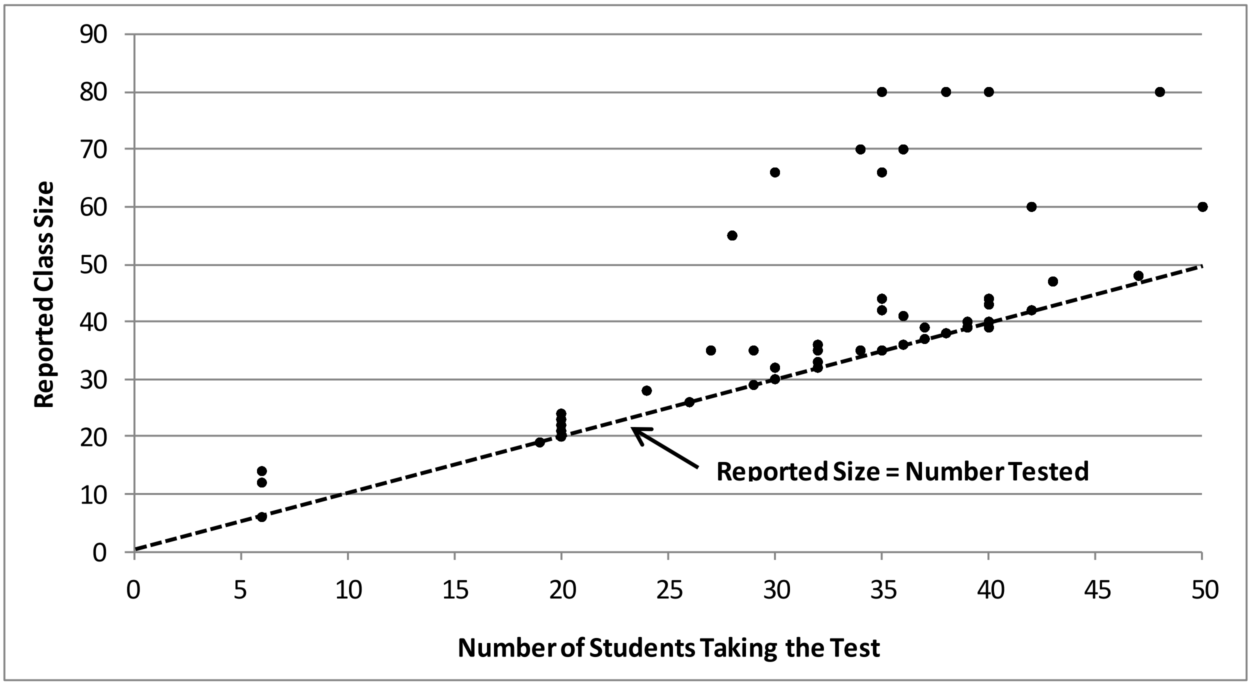
Table 1
Description of Data

Variable	**Unweighted Mean**	**Range**	**Observations**
Test Score (Imputed)	377	151-649	2361
Reported Class Size (Incorrect)	44.1	6-80	1769
Class Size as Tested*	35.6	3-50	2361
Reported Class Size (Revised)	36.7	6-60	1375
Female Gender	0.50	0 -1	2359
Computer at Home	0.46	0 -1	2239
>10 Books at Home	0.66	0 -1	2361
Share of Wealthy Students in School (0-10% to >50%)	1.60	1-4	1931
Math HW Frequency	2.65	1 – 3	2073
Teacher Education	4.83	2 – 6	2271
Teacher Experience	21.3	1 – 40	2261
Private School	0.19	0 – 1	2361
Rural School	0.12	0 – 1	2361
School Size	1846	36 – 6288	2166
Group by Ability (Yes = 1)	1.87	1-2	2337
*Students from the class present on the day of the test			

In a more in-depth review of the data, I determined that many teachers with two 4^{th} grade classes had included the total number of students in both classes as their estimate of the size of each class. These data are collected in question at4mstud on the teacher questionnaire, which asks: "How many students are in the TIMSS class for mathematics?" [IAEE, 2007] For a teacher with two 4^{th} grade classes, it is not clear if the question refers to the number of students in each class or in the two classes together. An additional problem with these data is that about 25% of the observations are missing. Although many variables have missing observations, this variable has the most missing observations, perhaps because teachers did not know how to interpret the question.

Figure 2

Reported Class Size vs. Class Size as Tested



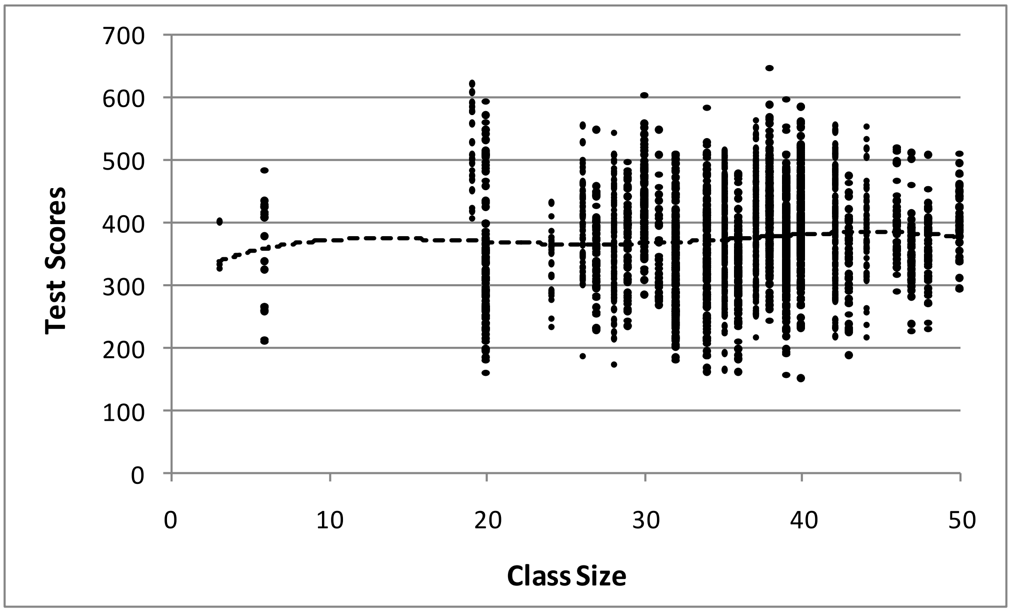
Due to these measurement problems, I decided to use the number of students taking the test as the primary estimate of class size in the study. The distribution of test scores as a function of this estimate of class size is shown in Figure 3. There is no apparent relationship between average test scores and average class size in these data.

The plot shows that the data include 21 students in very small classes (6 or less) and that they had lower average test scores. These students are in rural schools, which must be in isolated areas. I considered these data to be outliers and excluded them from the statistical regressions.

Use of the class size as tested rather than the reported class size could bias the estimated coefficients, since it likely incorporates two types of error into the data. First, the number taking the test in most cases is an underestimate of the number in the class. Second, the degree of error in each measurement depends on the absentee rate, which is random. The first error biases the class size estimate downward, which should bias the estimated effect of class size upward. The second error creates attenuation bias, which biases the estimated effect downward.

Figure 3

Test Scores vs. Class Size as Tested in the Data Set



But another possibility is that the number taking the test is a good estimate of the number normally present in the class, which is the number that should affect the test scores. If so, then the number tested is a more accurate estimate of the true class size affecting student achievement than the official class size. In this case the estimated coefficient using the number tested would have less attenuation bias than the estimated coefficient using the (correctly-reported) official class size. .

I created a smaller data set using the teachers' estimates of class size by identifying the incorrect estimates of class size and eliminating these observations from the data set. An examination of the data in Figure 2 reveals that the largest number of students taking the test was 50, while the reported number of students in that class was

60. In other classes the difference between the reported class size and those taking the test was either less than ten or much more than ten. This pattern indicated that a data set with valid class size data could be created by eliminating any data for reported class size that exceeded the number of students taking the test by more than ten.

This sort of the data reduced the observations in the reported class size data from 1769 to 1375, leaving only 78% of the original observations and providing data for 58% of the students tested. In this subset of data, the average class size was 36.7, of which 34.5 students, or 94.0%, took the TIMSS test. While this data set is considerably smaller than the data set using the number tested, I estimate the model using these data to provide a robustness check for the results using the larger data set.

The effect of class size on test scores is expected to be negative, but it is not clear *a priori* whether this effect is linear or non-linear. I examined the effect on test scores of both class size and $(\text{class size})^2$ to determine which provided superior statistical estimates.

2.2. Other Characteristics

The only personal characteristic included in the model is the student´s gender. Gender has been shown to affect scores in some tests of mathematics skills. The ICFES [2010] report shows that female students in Colombia have lower scores on the TIMSS 2007 mathematics tests than males.

Many studies of student achievement have found a large effect from students´ family characteristics [Hattie, 2009]. Both family income and parents´ education generally have a positive effect on student test scores. I included three variables to control for family characteristics: 1) whether the student has a computer at home, 2) how many books are in the home, and 3) the share of wealthy families in the school. The estimated coefficients on all of these variables are expected to be positive. I also tested whether the share of disadvantaged families in the school explained any test score variation and found that it had little effect.

According to the ICFES [2010] report, test scores in mathematics are highly correlated with the number of books in the home in 8th grade, but not in 4th grade.[2] I confirmed that the number of books across the four size categories does not explain the variation in scores in 4th grade and then collapsed the data into two categories, 0-10 books, and more than 10 books. The number of books in these two categories had some explanatory power.

Studies normally include data on teacher education and experience, since in theory these characteristics should have a positive effect on student achievement. I included variables for both characteristics, even though Glewwe, et. al. [2011] report that empirical studies usually find little or no effect. Teacher education is measured in levels from 2 to 6, but most teachers in the data set are in level 4. Experience in the data set is measured in years.

Studies sometimes find that the frequency and/or the amount of homework assigned have positive effects on student achievement [Glewwe, et. al., 2011]. TIMSS 2007 has both student and teacher questions related to this factor. I included the homework frequency variable from the teacher questionnaire in some models because the coefficient on this variable had the correct sign. In these data the frequency ranges from homework for some lessons (1) to half the lessons (2) to all or almost all lessons (3).

For school characteristics I included the size of the school, whether the school is urban or rural, and whether the school is private or public. Juerges and Schneider [2004] found that students in larger schools have higher scores. The ICFES [2010] report shows that on average 4th grade mathematics students in Colombia score higher in private schools than in public schools and higher in urban schools than in rural schools. These variables for the location and the type of school implicitly may provide additional controls for family characteristics, since family income is likely to vary in a consistent manner between these school categories.

[2] Woessmann [2005] found that the number of books in the home is strongly correlated with student performance in PIRL 2001 literacy tests in 4th grade, but he observed that the effect in Colombia is small because most Colombian households have few books.

2.3 Potential Bias and Limitations of the Study

Many estimates of the effect of class size on student achievement exhibit endogeneity bias. Lindahl [2005], Woessmann and West [2006], and Jakubowki and Sakowski [2006] show that OLS estimates of the effect of class size on student test scores in OECD countries are biased because schools often place students with learning problems in smaller classes. This practice causes OLS estimates of the effect of class size test to be positive, while estimates of the effect that include controls for this bias are negative.

The TIMSS 2007 questionnaire for School Directors included a question about whether students are grouped by ability. The data indicate that 300 students attended schools that group by ability. I removed these students from the data set in most of the regressions to control for endogeneity bias. Any remaining endogeneity should bias downward the estimated negative effect of class size on test scores.

The data set had numerous missing observations, and the number missing was different for each variable. As a result, the number of students included in each regression depends on the particular variables included in the model. This number is reported for each regression. All of the regressions had samples considerably smaller than the 2361 students tested. The variations in the sample size make the estimates in the various models less comparable. The reductions in the sample size make the statistical results less representative of the Colombian population. This is a limitation of the study. Some studies attempt to estimate the value of the missing observations, but given that the TIMSS test scores are estimates of plausible values created using five different (and unknown) sets of characteristics for the student and the learning environment, it was not evident how the missing observations could be reliably estimated.

Another limitation of the study is that it does not control for students' mathematics skills prior to entry into 4^{th} grade. These data are not collected in TIMSS or in most other international evaluations. As a consequence, it is not clear if the estimated effect of class size is due primarily to the student´s experience in 4^{th} grade or to the cumulative experience during the student´s first four years of primary school. Since most 4th grade students are likely to have been in the same school in a similar size class since 1st grade, the estimated effect of class size on test scores in this study is most likely the cumulative effect of this class size during the first four grades.

III. Empirical Results

Table 2 presents the empirical results using the number of students tested as the estimate of class size. Column 1 presents the results for the complete model, which has 1357 observations. In this data set the effect of class size is negative and quite large, but it is not significant at the 5% level. Each additional student in a class reduces student scores by 2.5 points.

Column 2 presents the results for the same model, excluding the students in schools that group by ability. This restriction reduces the sample size to 1247 observations and reduces the magnitude and significance of the effect of class size. It appears that in contrast to the OECD countries, grouping by ability in Colombia may assign the more able students to smaller classes. In any event all of the subsequent regressions in Tables 2 and 3 exclude the students in schools that group by ability.

Since teacher education and experience have virtually no effect on test scores in the first two regressions, these variables are eliminated in column 3. With this change the effect of class size becomes larger and significant at the 5% level. Each additional student reduces test scores by 3 points. But since school size and class size are correlated and the estimated coefficients have opposite signs, the magnitude of the coefficients on these two variables may be too large.

Column 4 examines whether the effect of school size is robust to elimination of the class size variable from the model. The resulting smaller and insignificant estimated coefficient on school size indicates that school size may not actually have much effect on test scores. Column 5 shows the results without the school size variable, and again the estimated coefficient on class size is smaller than when both school size and class size are included in the model. The estimate of the effect of class size excluding school size seems more reliable, and it is still statistically significant at the 5% level. In this model each additional student reduces test scores by 2.7 points.

Table 2
Effect of Class Size as Tested on Student Test Scores

	1	2	3	4	5	6	7	8
Observations	1357	1247	1287	1287	1442	1442	1442	1442
Female Gender	-14.2* (5.9)	-11.0 (6.4)	-10.1 (5.8)	-12.1 (6.4)	-10.5 (5.5)	-10.7 (5.6)	-10.7 (5.5)	-10.9 (5.6)
Computer at Home	9.1* (5.2)	10.6 (6.0)	12.5 (5.6)	17.2 (7.0)	11.5 (5.3)	11.6 (5.2)	12.1 (6.0)	12.4 (6.0)
>10 Books at Home	11.5 (5.2)	8.8 (5.5)	8.8 (5.7)	10.1 (6.1)	8.6 (5.2)	8.9 (5.3)	9.2 (5.2)	9.6 (5.3)
Wealthy Share of School	16.4 (9.4)	20.1 (11.9)	13.6 (5.6)	16.9 (8.4)	13.9 (5.6)	15.1 (5.8)	15.9 (6.0)	17.1 (6.4)
Math HW Frequency	23.9 (24.9)	20.1 (24.7)	15.8 (25.6)	20.6 (24.9)	21.1 (19.4)	20.8 (19.3)	22.5 (19.2)	22.4 (19.0)
Teacher Education	-0.4 14.2	0.7 (14.5)						
Teacher Experience	-0.23 (.79)	-0.06 (.87)						
Private School	8.0 (27.6)	4.5 (24.2)	22.6 (17.7)	10.6 (18.4)	11.0 (12.6)	12.4 (13.1)		
Rural School	-82.8* (21.0)	-99.0* (30.4)	-102.8* (24.3)	-64.3* (21.5)	-112.3* (21.1)	-102.7* (18.5)	-113.1* (21.7)	-103.7* (18.5)
School Size	0.003 (.006)	0.005 (.007)	0.009 (.006)	0.005 (.006)				
Class Size as Tested	-2.49 (1.59)	-2.02 (2.03)	-3.00 (1.24)		-2.68 (1.12)		-2.55 (1.15)	
Class Size^2 (as Tested)						-0.037 (.015)		-0.035 (.015)
Constant	386.0 (142.8)	359.5* (157.5)	401.9* (91.5)	286.4* (75.6)	393.5* (73.2)	344.7* (60.9)	385.1* (72.4)	337.3* (60.1)
R^2	0.24	0.24	0.22	0.18	0.21	0.21	0.21	0.21

Note: Robust standard errors in parentheses
*Statistically significant at 1% level

Column 6 tests whether the effect of class size on test scores may be non-linear. The statistical results indicate that $(\text{class size})^2$ is slightly more statistically significant than class size, although either variable provides similar estimates of the class size effect. In these results the estimated coefficient on private school is small and not statistically significant. Columns 7 and 8 show the results with the variable for private school removed from the model. The estimated coefficients on the class size variables in these models are slightly smaller but still statistically significant at the 5% level. These estimates provide a conservative estimate of the effect of class size on test scores. Over the range of class sizes as tested (19 to 50), these estimates indicate that the increase in class size reduces test scores by 75-79 points.

Overall the results in this table provide strong evidence that 1) family income raises test scores, 2) large classes and rural settings reduce them, and 3) teacher characteristics have no effect. These results are consistent with the other studies of the determinants of student achievement in 4^{th} grade. The magnitude of the explained variation in test scores (0.21 to 0.24) is consistent with other national studies using international test score data [Fuchs and Woessmann, 2007]

Table 3 presents the results of the model using the reported class size provided by the teachers. Column 1 shows the results using all of the estimates, including those that appear to be in error. The effect of class size is negative, but the effect is small and not statistically significant.

Column 2 presents the results for the same model but using only the observations for reported class size that do not exceed the number of students tested by more than ten. The estimated coefficient on class size is -2.04, and it is statistically significant at the 10% level. Columns 3 presents the results for $(\text{class size})^2$, and columns 4 and 5 present estimates for these two variables for class size without the variable for private schools. The empirical results using these data are all similar to the results in Table 2, except that they are slightly smaller and slightly less statistically significant. Over the range of reported class sizes (19 to 60), the increase in class size reduces test scores by 78-83 points using the two class size variables. Since the estimates of class size are higher in the reported class size data, the net effect of changes in class size on tests scores in these estimates are 4-5% larger than the effect using the class size data based on the number tested.

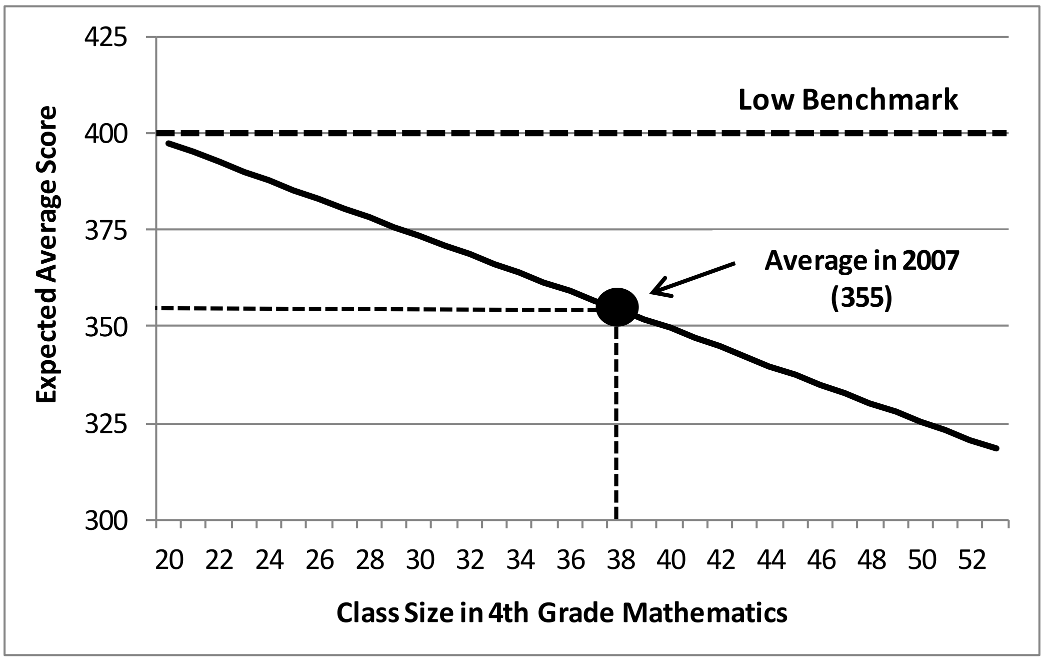
Table 3 Effect of Reported Class Size on Student Test Scores					
	1	**2**	**3**	**4**	**5**
Observations	1228	962	962	962	962
Female Gender	-11.4 (6.6)	-8.6 (7.2)	-8.6 (7.3)	-8.6 (7.1)	-8.6 (7.2)
Computer at Home	14.1 (6.6)	15.8 (6.8)	16.2 (6.8)	16.8 (8.0)	17.4 (8.1)
>10 Books at Home	11.9 (6.1)	10.6 (5.9)	10.9 (6.0)	11.2 (5.8)	11.6 (6.0)
Wealthy Share of School	15.8 (8.48)	15.3 (7.8)	15.7 (8.1)	17.5 (7.9)	18.5 (8.4)
Math HW Frequency	18.4 (25.8)	38.4 (16.9)	38.9 (16.2)	39.3 (16.4)	40.0 (15.5)
Private School	3.8 (14.4)	12.7 (13.9)	13.7 (14.3)		
Rural School	-82.8* (20.7)	-101.0 (20.3)	-92.3* (17.5)	-104.1* (19.8)	-95.5 (16.4)
Class Size (Bad TIMSS data)	-0.65 (0.76)				
Class Size (TIMSS data)		-2.04 (1.15)		-2.03 (1.20)	
Class Size ^2 (TIMSS data)			-0.024 (.013)		-0.024 (.014)
Constant	329.1* (74.5)	316.9* (72.6)	272.0* (51.4)	314.1* (73.6)	268.6* (50.7)
R^2	0.19	0.27	0.27	0.26	0.26
Note: Robust standard errors in parentheses *Statistically significant at 1% level					

Even though the data samples are different for the two measures of class size (1442 vs. 962 observations), these two sets of results are consistent. This similarity provides reassurance that the effect of class size on TIMSS test scores is robust.

I estimated earlier that the absentee rate on the day of the test was 6%. Applying this factor to the number taking the test converts the average of 35.6 and the range of 19-50 students taking the test to an average of 37.8 and a range of 20-53 students actually in the classes. Figure 4 shows the estimated relationship between expected test scores and actual class size in Colombia after controlling for other factors. The relationship is calculated from the estimate of the effect of class size on test scores in column 7 in Table 2. This estimate of the effect is conservative, since the estimate in column 5 is larger. After adjusting for the absentee rate, each additional student reduces test scores by 2.4 points (2.55/1.06). An increase in class size from 20 to 53 reduces average test scores by 79 points.

Figure 4

Expected Test Score vs. Class Size in Colombia (Conditional)



These results provide evidence that a reduction in primary school class sizes could be an effective strategy to raise Colombian student achievement in mathematics. According to the study, reducing the size of all mathematics classes in primary school to 20 students would increase average test scores in Colombia from 355 to 398.

These results can be used to estimate the effect size of specific reductions in class size. The standard deviation in test scores in the sample is 79. The implication is that a reduction in class size from 45 to 25 would have an effect size of 0.61. An effect size of this magnitude is considered a large effect [Hattie, 2009].

The results also identify another policy change that could raise student test scores. The two sets of results both obtain a positive effect from an increase in the frequency of mathematics homework. In the results using the number of students tested, an increase in homework frequency from half the lessons to all the lessons increases scores by 22 points, but the estimate is not statistically significant. In the results using the reported class size, this change increases student test scores by 40 points, and the estimate is statistically significant at the 5% level. The data indicate that 65% of the 4th grade students in the sample already have homework for all lessons, but that still leaves 35% who could increase their scores if they were given homework more frequently.

IV. Conclusions

This study examines whether large classes in primary school in Colombia have an adverse effect on student scores on tests of mathematics skills in 4th grade. After controlling for numerous characteristics of the students' learning environment, the results indicate that each additional student in a class is associated with a decline in average test scores of 2.4 points. Applying the estimated coefficients to the data in the sample indicates that an increase in class size from 20 to 53 students is associated with a decline in test scores of about 80 points. This change is very large compared to the low benchmark of acceptable skills of 400 and the imputed average score in Colombia of 355. It is also large relative to the standard deviation of the test scores in the sample, which is 79.

Although it is not evident from the analytical results, it seems likely that the estimated effects of smaller classes on test scores are the result of smaller classes throughout the first four years of primary school and are <u>not</u> due only to differences in 4th grade. It is not clear what the effect would be if class sizes were reduced only in 4th grade. Longitudinal studies would be required to determine the effects of reductions in class size on achievement in the different primary grades.

These analytical results depend crucially on the selection of the class size data used in the analysis. An analysis of these data revealed that the question about class size in TIMSS 2007 is not well-specified, and as a result, for 22% of the observations, teachers apparently provided an estimate of the total number of students they teach in 4th grade rather than the number in each class.

As a consequence, if the effect of class size on test scores is estimated using all of the reported data, the estimated effect is negative, but it is small and not statistically significant. The empirical results only show that class size has a large, statistically-significant negative effect on test scores if the misreported data are excluded or if the number of students tested in each class is used as the estimate of class size.

The adverse effects of class size found in this study are larger and more statistically significant than the effects found using TIMSS data in many other countries. There are several possible explanations for this difference. The most likely is that class sizes in Colombia are larger on average and exhibit more variation than in most OECD countries. But it is possible that the TIMSS data on class size are also misreported in other countries. The results here indicate that if researchers have not identified and eliminated these misreported data, they would not find that class size has a large adverse effect on student achievement.

Reducing class size is expensive. Sanchez [2006] estimates that teachers' salaries constitute 85-90% of the cost of schooling in Colombia, so unit schooling costs are much higher in small classes. But the focus here is only on mathematics classes, not all primary school classes. The size of mathematics classes in primary school could be reduced at a more reasonable cost by providing additional teachers only for mathematics classes.

Metzler and Woessmann [2010] show that greater teacher competence in mathematics (as measured on tests) substantially raised test scores in 6th grade mathematics in Peru. Their results suggest that if the additional teachers provided to reduce class size have high competence in mathematics, student scores might improve substantially.[3] Perhaps some test schools could be selected for a controlled experiment

[3] The regression results show that teacher education and experience do not affect test scores in 4th grade mathematics, but they do not indicate whether teacher competence in mathematics affect scores.

to evaluate whether such an approach is effective in raising Colombian student achievement.

The empirical results indicate that test scores for students in 4th grade mathematics may be higher in larger schools. It may be that larger schools already use specialized teachers for mathematics more frequently than smaller schools. If this is the case and they obtain superior results, then the increased use of such teachers could be a cost-effective method for raising test scores.

References

Breton, Theodore R., 2011, "The Quality vs. the Quantity of Schooling: What Drives Economic Growth?," *Economics of Education Review*, v30, 765-773

Foy, Pierre, and Olson, John F., *TIMSS 2007 User Guide for the International Data Base*, TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College

Fuchs, Thomas, and Woessmann, Ludger, 2007, "What Accounts for International Differences in Student Performance? A Re-examination Using PISA Data, *Empirical Economics*, v32, 433-464

Glewwe, Paul W., Hanushek, Eric A., Humpage, Sarah D., and Ravina, Renato, 2011, "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010," NBER Working Paper 17554

Hanushek, Eric A., and Luque, J.A., 2003, "Efficiency and equity in schools around the world," *Economics of Education Review*, v22, n4, 481-502

Hanushek, Eric A. and Woessmann, Ludger, 2008, "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46.3, 607-668

Hattie, John, 2005, "The Paradox of Reducing Class Size and Improving Learning Outcomes, *International Journal of Educational Research*, v43, 387-425

Hattie, John, 2009, *Visible Learning*, Routledge, New York

IAEE (International Association for the Evaluation of Education Achievement), 2007, *Student, Teacher, and School Questionnaires*, TIMSS 2007, http://timss.bc.edu/timss2007/PDF/T07_Questionnaires_G4.pdf

ICFES, 2010, *Resultados de Colombia en TIMSS 2007*, ICFES Evaluaciones Internacionales, Bogota, D.C.

Jakubowski, Maciej, and Sakowski, Pawel, 2006, "Quasi-experimental Estimates of Class Size Effect in Primary Schools in Poland, *International Journal of Educational Research*, v45, 202-215

Juerges, Hendrik, and Schneider, Kerstin, 2004, "International Differences in Student Achievement: An Economic Perspective," *German Economic Review*, v5, n3, 357-380

Lindahl, Mikael, 2005, "Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement," *The Scandinavian Journal of Economics*, v107, n2, 375-394

Macdonald, Kevin, 2008, *PV: Stata module to perform estimation with plausible values,* Statistical Software Components S456951, Boston College, Department of Economics, revised 12 Apr 2011, http://fmwww.bc.edu/repec/bocode/p/pv.ado

Metzler, Johannes, and Woessmann, Ludger, 2010, "The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation," CESIFO Working Paper No. 3111

Mullis, I.V.S., Martin, M.O., and Foy, P., 2008, *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, TIMSS and PIRLS International Study Center, Boston College, http://timss.bc.edu/timss2007/intl_reports.html

Olson, J.F., Martin, M.O., and Mullis, I.V.S., 2008, *TIMSS 2007 Technical Report*, TIMSS and PIRLS International Study Center, Boston College, http://timss.bc.edu/timss2007/techreport.html

Parcel, Toby L., and Dufur, Mikaela, 2009, "Family and School Capital Explaining Regional Variation in Math and Reading Achievement," *Research in Social Stratification and Mobility*, v27, 157-176

Rivkin, Steven G., Hanushek, Eric A., and Kain, John F., 2005, "Teachers, Schools, and Academic Achievement, *Econometrica*, v73, n2, 417-458

Sánchez, Gustavo Alberto, 2006*, "Análisis Costo Promedio por Alumno en Bogotá, D.C.," Ensayos Sobre Política Económica*, Banco de la Republica, Revista No. 51, Edición Especial Educación.

Woessmann, Ludger, 2005, "Families, Schools, and Primary-School Learning: Evidence for Argentina and Colombia in an International Perspective," World Bank Policy Research Working Paper 3537

Woessmann, Ludger, and West, Martin, 2006, "Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS," *European Economic Review*, v50, 695-736