International Journal of Educational Development 34 (2014) 51–57

Contents lists available at SciVerse ScienceDirect

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev

Evidence that class size matters in 4th grade mathematics: An analysis of TIMSS 2007 data for Colombia

Theodore R. Breton *

Universidad EAFIT, Carrera 49 #7 Sur-50, Avenida Las Vegas, Medellin, Colombia

ARTICLE INFO

Keywords:
International education
Colombia
Class size
Test scores
TIMSS

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.

1. 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; 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).

Fig. 1 shows the average test scores in mathematics in 4th and 8th grade for a subset of the countries that participated in TIMSS 2007 (Mullis et al., 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).

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% 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

* Tel.: +574 261 9334; fax: +574 261 9294.
E-mail addresses: tbreton@eafit.edu.co, ted.breton@gmail.com

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

0738-0593/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijedudev.2013.04.003

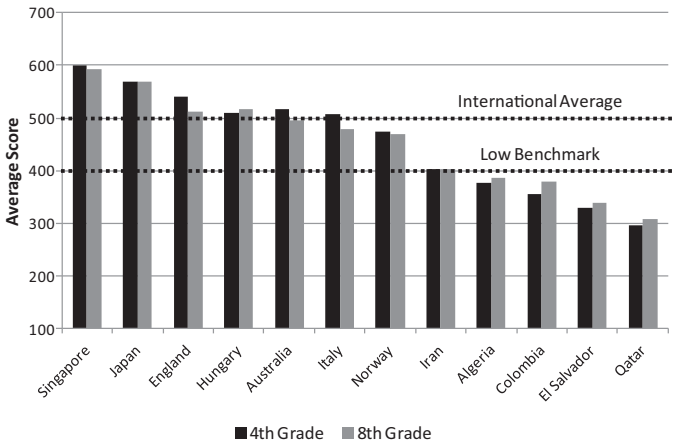

Fig. 1. Average scores on TIMSS 2007 tests of mathematics skills.

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 meta-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 et al. (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 nine but not at age 13.

In a very recent meta-analysis of studies performed in developing countries since 1990, Glewwe et al. (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 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 effect in 4th grade likely measures 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 was misreported. Many teachers reported the total number of 4th 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 2 presents the methodology used in the analysis. Section 3 presents the results. Section 4 concludes.

2. 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):

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

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 et al. (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.