Do longer school days improve student achievement? Evidence from Colombia

Diana Hincapié*

May 7th, 2014

Abstract

This paper analyzes the impact of longer school days on student achievement in Colombia, where primary and secondary students attend schools that have either a complete (7-hour) or a half-day (4-hour) schedule. Using test score data from 5th and 9th graders in 2002, 2005, and 2009, along with school administrative data, this study identifies the effect of longer school days by implementing a school fixed effects model. The main model compares variation in average test scores across cohorts for schools that switched from a complete schedule to a half schedule and vice versa. I find that among schools that switch schedules between 2002 and 2009, the cohorts exposed to complete schedules have test scores that are about one tenth of a standard deviation higher than cohorts that attended half schedules. The impact of a complete schedule is larger for math test scores than for language test scores, and it is larger for 9th grade test scores than for 5th grade test scores. Effects are largest among the poorest schools in the sample and those in rural areas. The results suggest that lengthening the school day may be an effective policy for increasing student achievement, particularly for the lowest-income students in Colombia and other developing countries.

Keywords: quality of education, student achievement, instructional time, school schedule.

JEL codes: I21, I28, H43.

---

* Trachtenberg School of Public Policy and Public Administration, The George Washington University. Email: dianah@gwmail.gwu.edu.

I want to thank my advisor, Stephanie Cellini, for her outstanding support, guidance, and mentoring. This paper has benefited from discussions with Dylan Conger, Burt Barnow, Felipe Barrera, Hal Wolman, Sandra García, Olga Romero, and participants at the Trachtenberg School of Public Policy and Public Administration's research seminar and the 18th annual meeting of the Latin American and the Caribbean Economic Association.

1. Introduction

There are large differences in the time students spend in school, both across and within countries. Are these variations in schooling contributing to student achievement gaps and education inequalities? It seems to be a widely held belief that the more time students spend in school, the more they will learn, and so the number of hours (length of the school day) and the number of school days (length of the school year) are directly and positively related to student learning and therefore, to student achievement. While there is a large literature that provides evidence about the effect of school inputs in the education production function, there is only limited evidence on the effects of instructional time on student outcomes, and even less evidence showing that interventions that increase the length of the school year or the school day will improve learning outcomes. Overall, research supports the idea that additional instructional time has a positive impact on student achievement and other student outcomes. [1] However, there are methodological difficulties in isolating the impact of instructional time that raise questions about the strength of much of this evidence.

This study contributes to the literature by analyzing the causal link between longer school days and student achievement in Colombia, using a school fixed effects identification strategy that mitigates some of the most critical selection and endogeneity problems found in the literature.

In Colombia, some schools offer a full 7-hour school day, known as a "complete" schedule, while others offer two separate 4 or 5 hour shifts, known as "half" schedules.[2] The additional time in the former is usually devoted to extra hours of instruction. Cross-sectional analyses comparing student test scores in schools with complete and half schedules will likely be

[1] I present a detailed review of the literature in section 3.
[2] In half schedule schools, two different groups of students attend the same school, with one group attending the morning shift, and the other one the afternoon shift.

biased since schools that have a complete schedule may have unobservable characteristics that are related to student achievement (e.g., more educated parents, or more motivated principals and teachers). However, in Colombia, schedules are determined at the municipal level according to policy and projected enrollment, and thousands of schools switched from half to complete (and vice versa) in the time period I study. I can therefore compare the test scores of cohorts of students in the same school before and after the switch to (or from) a complete schedule. This strategy can effectively control for all unobserved time-invariant characteristics of the school that may bias cross-sectional estimates, although time-varying unobservables may remain.[3]

I draw on data from the 2002, 2005, and 2009 SABER to implement this approach. The SABER is a nationwide standardized test administered every three years in all primary and secondary schools in Colombia. It evaluates 5th and 9th grade students' skills in both mathematics and language. I merge these data with administrative data on all primary and secondary schools in Colombia.

Policies aimed at increasing the length of the school day continue to be widely debated in Colombia, as well as in other countries. Arguing that extending the school schedule might have a positive impact on student outcomes, some countries, such as Chile and Uruguay, have implemented a complete schedule in all public schools that previously had a half schedule. Colombia implemented this same reform in 1994, but it was not enforced and was later rescinded, mainly because of the increasing demand for school spots that required having two half-day schedules in some schools. However, in recent years, some cities have starting piloting interventions to increase the length of the school schedule. As many countries and cities continue

[3] Family mobility may be a particular concern if students switch schools after the change in schedule. However, few families in Colombia move because of schools and fewer still move before their children reach the 5th and 9th grades. I address these concerns in section 4.

to experiment with this type of programs and policies, this study aims to contribute to the debate about the effectiveness of increasing the length of the school day.

This paper contributes to the literature on instructional time by analyzing the impact of time in school given by the length of the school day rather than the length of the school year, where a substantial portion of the literature has focused. It also helps fill the gap in the literature on the effect of instructional time in a developing country, and especially in Colombia, where evidence is particularly scarce.

To the best of my knowledge, this is one of only three studies that estimates the impact of longer school days in Colombia, and the first one to look at the impact on student achievement for 5th grade and 9th grade students in the country: Garcia, Fernandez, and Weiss (2012) look at the impact on dropout and grade repetition of 1st and 2nd grade students, while Bonilla-Mejia (2011) looks at the impact on student achievement in a high school exit exam (ICFES). This paper contributes to the literature on longer school days in Colombia in three main ways. First, because there might be different impacts of having longer schools days for students of different ages or grades, this study analyzes the impact on students attending different grades than those that have been studied before. Second, by analyzing 5th grade and 9th grade students in the same study, using the same type of data and methodology, this study allows comparing the differences in the magnitude and statistical significance of the impact of longer schools days between students attending primary and secondary grades. Third, by using a school fixed effects methodology that compares different cohorts of students within a school, this study is able to mitigate many of the selection concerns that are common when comparing differences between schools.

I find that among schools that switch schedules between 2002 and 2009, the cohorts exposed to complete schedules have test scores that are about one-tenth of a standard deviation higher than cohorts that attended half schedules. The impact of a complete schedule is larger for math test scores than for language test scores (e.g., 0.138 v. 0.110 of a standard deviation respectively, for 9th graders) and it is larger for 9th grade test scores than for 5th grade test scores (e.g., 0.138 v. 0.082 of a standard deviation respectively, for math test scores).

Using different subsamples to estimate the impact of a complete schedule, I find that the effect of attending a complete schedule is largest among rural schools and among the very poorest schools in the country. Overall, the results suggest that lengthening the school day may be an effective policy for increasing student achievement, particularly for 9th grade students, and for the lowest-income students in Colombia.

The rest of the paper is organized as follows: Section 2 describes the background on school schedules in Colombia. Section 3 provides a brief conceptual framework and reviews the most relevant literature on instructional time. Section 4 discusses the empirical strategies. Section 5 describes the data. Section 6 presents results, and Section 7 concludes.

2. Background

Over the last 20 years policymakers in Colombia have continually debated the issue of the appropriate length of the school day. In 1994, the government of Colombia passed the General Education Law (Law 115), which established that all public schools should have a single complete (7-hour) schedule. However, plans to implement this law did not begin until years later and were completely abandoned by 2002. This happened partly due to a substantial increase in demand for public schooling that was difficult to meet because of low physical capacity (fewer schools than the ones needed to meet the demand) and scarce resources (low capacity to hire enough administrators and teachers for a full day).

More recently, there have been proposals to establish a longer schedule in all public schools in the country, coming both from academia (Barrera-Osorio, Maldonado, & Rodriguez, 2012), and from policymakers, particularly in two of the largest cities in the country, Bogotá and Cali, where pilot programs to increase the length of the school day are being implemented. In April 2012, Bogotá city announced a pilot to turn 25 public schools into complete schedule schools, and Cali's Mayor announced that 2 schools in the city (encompassing approximately 1,000 students) will start having a complete schedule rather than the previous half-day schedules, and that this new standard will be implemented in the entire city during his term. In both cases, however, the additional time would be devoted to civics, sports, and cultural activities, rather than additional instructional time.[4]

The argument for implementing a longer school day is that more time spent in school will translate into better outcomes for students. However, except for a few recent studies that have found a positive impact on student achievement (Bonilla-Mejia, 2011) and other student outcomes (Garcia, Fernandez, and Weiss, 2012), there is not a lot of evidence suggesting a causal link between the length of the school day and student outcomes in Colombia.

Establishing the impact of having a complete schedule on student outcomes is methodologically difficult. If students were randomly assigned to different school schedules, in principle one could analyze the impact of attending a complete schedule by estimating the

[4] Although it is not possible to know what students are really doing during the additional hours in school from the data, in the current situation it is very likely that students attending schools with a complete schedule are receiving additional instructional time.

difference in student achievement between students attending a complete schedule and students attending a half-day schedule. However, a student's application and allocation to a school, and the decision process on which type of schedule a school has, is neither simple nor random.

First, the Secretary of Education of each municipality assesses the expected demand for school spots for the following year, and the capacity that schools in that municipality have to offer school spots (i.e., supply). Depending on that information, he or she determines whether it might be necessary to have two half-day schedules (a morning and an afternoon shift) instead of a complete schedule in certain schools, in order to increase the supply of school spots (Ministerio de Educación Nacional, 2006). As a result, the process of deciding the schedule for the municipality's schools is a function of many variables. These include the number of people who are at school age in the municipality, the actual demand for school spots (i.e.; the number of students already enrolled in the education system, and the applications filled out by parents for each school), the number of available places in each grade in each school, and the amount of education resources in the municipality (e.g., to hire more teachers). In some cases, the Secretary of Education might decide to implement half-day schedules in schools with a higher demand in order to be able to accept more students in those schools, and to implement complete schedules in schools with a lower demand. If the demand for each school is related to its quality (e.g., higher demand for relatively higher quality schools, and lower demand for relatively lower quality schools), this could possibly cause a downward bias in Ordinary Least Squares (OLS) estimates of the impact of a complete schedule, because schools with a complete schedule will have other unobserved characteristics related to lower test scores.

Second, in Colombia, parents cannot directly choose a public school for their children: When they want to enroll their children for the first time into a school, or want to transfer their children into a different school because they moved to a new area, they need to complete an application with the municipality's Secretary of Education. In this application, parents are able to express their preference for two schools, but they cannot express their preference for a particular schedule. The Secretary of Education then makes the decision on where to assign the new or transferred students whose parents have applied for a school spot, based on the demand and supply of schools spots for each grade. Ideally, this decision matches the parents' expressed preferences, but it is possible that children are assigned to a school that is not in their parents' preferences if there are not enough schools spots in the parents' preferred schools.[5]

3. Conceptual framework and literature review

i. Impact of the length of the school year

The literature on the impact of instructional time can be classified in two main categories: One that analyzes the impact of differences in the length of the school year, and one that looks at the impact of longer school days or more instructional time in a specific subject per day.

Most of the literature has focused on analyzing the impact of the school year's length on student achievement or other outcomes. For example, analyzing the effect of schooling on labor market outcomes, Margo (1994) found evidence that historical differences in the length of the school year accounted for a large fraction of the differences in earnings between black and white workers at the end of the 20th century. Lee and Barro (2001) compared differences in the length of school year across countries and found that more time in school improved math and science

[5] If that is the case, the municipality will try to allocate the students to the school closest to their place of residence. However, it might be possible that they are assigned to a school in the other side of the city if that is the only school where there are available spots, although this is not frequently the case.

test scores, but it seemed to lower reading scores. Eren and Milliment (2007) studied the variation in the length of the school year and the average duration of classes across states in the United States. The results indicated that the school year length, and the number and average duration of classes, affected student achievement. However, the direction and magnitude of the effects were not homogeneous across the test score distribution: test scores in the upper tail of the distribution benefitted from a shorter school year, while a longer school year increased test scores in the lower tail.

Some authors have taken advantage of natural experiments to analyze the impact of time spent in school and student achievement. For example, Marcotte (2007) and Hansen (2008) exploited year-to-year differences in the length of the school year that occur as a result of the weather. They compared how schools in two states in the United States performed in years when there were frequent cancellations due to snow days, to how they performed in years when the winter was relatively mild. They found that an additional 10 days of instruction resulted in an increase in student performance on state math tests of roughly 0.2 standard deviations.

Related to the analysis of the length of the school year, a strand of the literature has analyzed the *summer learning loss*. This concept is based on the idea that the summer break might have a negative impact on student learning, because of a lack of cognitive stimulation, a break in the rhythm of instruction, and forgetting that requires a significant amount of review when students return for the start of the academic year (Cooper et al., 2010). For example, a meta-analysis of the impact of summer vacation on test scores found that summer learning loss was equivalent to at least one month of instruction (Cooper et al., 1996). An alternative to mitigate the impact of the summer learning loss is extending the length of the school year, or switching to year-round calendars with the same number of instructional days but spreading the days of instruction evenly across the year. The impact of this type of interventions is not clear. For example, McMullen and Rouse (2007) found that for Wake County, North Carolina, a switch to a year-round calendar had essentially no impact on academic achievement. And for Colombia, Zuluaga (2013) estimated that 10th and 11th grade students exposed to an increase in school breaks had on average 0.22 standard deviations lower test scores.

ii. Impact of the length of the school day or more instructional time per day

Most of the literature analyzing the length of the school day has focused on developed countries, and particularly in the United States. A large proportion of these studies have looked at the impact of extending the length of kindergarten. In a meta-analysis of the literature, Cooper, Batts Allen, Patall, and Dent (2010) found that attending a full-day kindergarten was positively associated with academic achievement, compared to a half-day kindergarten, but the effect generally seemed to fade-out by third grade. They also found a significant correlation between attending a full day and the child's self-confidence and ability to work and play with others. However, they found that children attending a full-day kindergarten did not have as positive an attitude toward school, and had more behavioral problems. Rathburn (2010) explored the relationships between full-day kindergarten program factors and public school children's gains in reading test scores. She found evidence that children who attended kindergarten programs that devoted a larger portion of the school day to academic instruction (particularly reading instruction) made greater gains in reading over the school year than children who spend less time in such instruction. However, the effects of a full-day kindergarten seem to fade out beyond the kindergarten year, something that might be related to summer learning loss (Redd et al., 2012).

Although there is a large literature on the impact of school time in early education programs, little is known about the impact on K-12 education. This is partly due to the small variation in the length of the school day across schools districts (in the US), or municipalities (in other countries), or to variations in length that are directly related to resources, which makes it difficult to estimate the impact of longer school days or longer school years. Some of the studies that have used identification strategies that allow them to establish a causal impact of school time on student outcomes have shown that spending more time in school could improve academic achievement and could benefits students, families and societies by reducing teen pregnancies (Kruger & Berthelon, 2009) and crime rates (Jacob & Lefgren, 2003).

There are a few papers that have analyzed the impact on test scores of variations in the length of instructional time across countries. Using data from the 2006 Programme for International Student Assessment (PISA)[6] for over 50 countries, Lavy (2010) found evidence that instructional time consistently had a positive and significant effect on test scores both in developed and developing countries. However, the estimated effect for developing countries was much lower than the effect size in developed countries. More recently, Rivkin and Schiman (2013) looked at the effect of instructional time on the 2009 PISA test scores results for 72 countries. They also found similar that achievement increases with instructional time, and that the increase varies by the amount of time and classroom environment.

In Latin America, some studies have exploited natural experiments given by the implementation of a full-day schedule in schools that previously had half-day schedules. Valenzuela (2005) did one of the first impact evaluations of the Chilean full school day program.

[6] The PISA is an international study jointly developed by the Organization for Economic Cooperation and Development (OECD) countries to evaluate education systems worldwide. Its main component comprehends testing the mathematics, science, and reading skills and knowledge of 15-year-old students in schools in participating countries or economies.

Implemented in 1997, the goal of this program was to increase by 30% the number of daily hours in public and voucher schools in the country. Taking advantage of the differences in the time in which schools implemented the program, the author implemented a Difference-in-Difference model at the school and county levels. He found that the program had a significant, positive, and robust effect on schooling outcomes, but the impact was different depending on socioeconomic characteristics of the treated schools and also among subjects. At the county level, the results also showed a positive impact of the program.

Also analyzing the Chilean full-day schedule program, Bellei (2009) found that it had positive effects on students' academic achievement in both mathematics and language, and these effects were constant over time. He also suggested that the program had larger positive effects on rural students, students who attended public schools, and students in the upper part of the achievement distribution. Pires and Urzua (2011) analyzed the effect of time spent in school on schooling attainment, cognitive test scores, socio-emotional variables, labor market outcomes, and social behavior in Chile. They concluded that enrollment into a full-day school had positive effects on academic outcomes and cognitive test scores, and reduced adolescent motherhood, but there was no effect on employment or wages. Cerdan-Infantes and Vermeersch (2007) analyzed the impact of a program that lengthened the school day from a half day to a full day in poor urban schools in Uruguay. They found that students in very disadvantaged schools improved test scores by 0.07 of a standard deviation in math, and 0.04 of a standard deviation in language.

Llach et al. (2009) analyzed the long-term impact of attending longer school days in Buenos Aires, Argentina. They found that students that attended full-day primary schools had a secondary school graduation rate 21 percent higher than those that attended half-day primary schools, and this was mostly driven by students coming from low socioeconomic backgrounds. However, they did not find enduring effects on income and employment.

For Colombia, Camacho and Mejia (2013) analyzed if the conditionality on school attendance imposed by a Conditional Cash Transfer program *(Familias en Acción),* lead to reductions in crime by impeding teenagers from getting involved in criminal activities. They found that the program does not seem to have an effect on crime through this "incapacitation" mechanism cause by school attendance.

To the best of my knowledge, there are only two studies that have evaluated the impact of longer schools days on student outcomes in Colombia. In the first one, Bonilla-Mejia (2011) evaluated the impact of attending a complete schedule school on student performance in a high school exit exam (SABER 11, previously known as "ICFES"). He used the number of students enrolled in complete schedule schools to instrument the probability that a student has of attending a complete schedule school. He found that students attending complete schedules have average test scores that are 2.5 percentage points larger than those attending half-day schedules. Additionally, they found that the impact was larger when compared with students attending only an afternoon shift. In the second study, Garcia et al. (2012) used family fixed effects to analyze the impact of changing from a half-day to a complete schedule on student outcomes. They found that a complete schedule reduces the probability of early dropout and grade repetition.

iii. Heterogeneous effects

The literature has found that the impact of more instructional time on student outcomes tends to be larger for students from poorer backgrounds or the most at-risk children. For example, Olsen and Zigler (1989) found that in general, extended-day programs in kindergarten seemed to increase standardized test scores in the short term, particularly for disadvantaged, bilingual, or “least-ready” for school children. Cooper et al. (2010) compared evaluations of full-day kindergarten programs in urban versus nonurban settings, and concluded that these programs had a significantly stronger association with higher academic achievement for children attending in an urban setting than those in non-urban communities. They suggested that children attending in urban settings were more likely to come from poorer backgrounds, so this could be taken as indirect evidence of a potentially greater impact of full-day kindergarten for poorer children. Harn, Linan-Thompson, and Roberts (2008) analyzed the impact of intensifying instructional time on early literacy skills for the most at-risk first graders. They concluded that the students receiving more intensive interventions (more instructional time) made significantly more progress across a range of early reading measures.

Additionally, there might also be heterogeneous effects on different types of students, because of variations across schools in the quality of instruction and of the content taught during the extra hours. For example, Llach et al. (2009) show that for Buenos Aires, the content of the additional hours was probably more important for academic achievement than increasing the number of hours. The impact on academic results could be very different depending on whether the extra hours are just an extension of the current curriculum, or whether they allow the disadvantaged students to develop their skills and abilities through instruction in the areas in which their more advantaged schoolmates usually learn and practice.

iv. Hypotheses

Given this framework, the impact of longer school days on student achievement in Colombia is uncertain. There are three arguments for why longer schools days could benefit students: First, by spending more time in school they will devote more time to learning. Second, they will spend less time alone at home or outside home doing educationally unproductive activities. And third, they will spend less time in the streets, exposing themselves to different risks (Kruger & Berthelon, 2009; Jacob & Lefgren, 2003). Therefore, having a complete schedule might have a positive impact on student achievement if the quality of the additional hours is good (i.e. stimulating or educationally productive), because students might be able to learn more things and improve their skills, and this could be reflected in higher student achievement.

On the other hand, longer school days could have a negative or null impact on student achievement if instruction during those extra hours is of poor quality. For example, if less prepared teachers are brought to schools in order to be able to extend the number of school hours, then the additional time spent in school will not necessarily translate into more learning for the students attending complete schedules. Alternatively, it may be the case that higher-quality schools have greater demand and are therefore “split” into a morning and afternoon shift, as mentioned previously. In this case, schools with complete schedules might then have lower test scores than half-schedule schools.

Additionally, the length of the school day could have no impact if the content of the extra hours in the complete schedule schools is largely irrelevant. For example, if they devote the additional time to sports, or more playing time, then there might not be a significant difference in student achievement between schools or cohorts that have a complete schedule and those that have a half-day schedule. Finally, the additional hours could have no impact if those attending a half-day schedule systematically come from better-off families or they can make up for the fewer hours spent in school.

The impact of longer school days might be heterogeneous. First, it might be larger for the poorest or most vulnerable children. The reason for this is that, assuming the quality of the additional hours of schooling is the same across all students and schools, they will benefit by more because that extra time in school will be relatively better for them than the home environment, compared to better-off children. For example, if not in school for those additional hours, students from higher-income families might be attending extra-curricular courses, playing with their parents, doing homework, or reading books. In contrast, students from lower-income families might spend relatively more time outside of school and outside their homes, increasing their exposure to risk factors associated with crime, teen pregnancies or drugs. The impact of longer school days might also be larger for the older students (i.e. 9th grade), because they are more likely to gain more from not being exposed to the risk factors associated with being outside the home, compared to the younger students, who are more likely to be at home if not in school for those extra hours. For example, older students who are not in school would most likely be in the streets, exposing themselves to crime and teen pregnancies, while younger students would most likely be at home or somewhere else, but in general less exposed to those risks.

4. Empirical strategy

The main goal of this study is to analyze the impact of longer school days (i.e., complete schedule) on student achievement. As described in section 2, there are methodological difficulties in isolating this impact, because there might be systematic selection of students to schools with a particular type of schedule, or of schools with particular characteristics into a type of schedule.

The following sub-sections describe several estimation strategies and the extent to which they can address endogeneity and selection concerns. The dependent variable is the average SABER test score for each grade, subject, and year for each school. I standardize the schools' test scores by grade, subject, and year, so that each of these samples of school test scores has a mean of zero and a standard deviation of one. This allows me to interpret the coefficients in the regressions as standard deviations from the mean. The key independent variable in this study is an indicator for the type of schedule that a school has in each year: COMPLETE[7] is equal to one if the school has a complete schedule in a year, or zero if it does not. Not having a complete schedule is equivalent to saying that the school has a half schedule (morning OR afternoon schedule) or double shifts (morning AND afternoon schedules).[8]

i. Estimation Strategy 1: Cross-sectional OLS

I first run a traditional cross-sectional OLS model. This model compares schools that have a complete schedule with those that have a half-day schedule in year *t*, controlling for observable characteristics and for school performance in the previous period. The use of previous period test scores requires dropping data for 2002 (because this year does not have a previous period test score). Therefore, this model uses a sample which contains only data for 2005 and 2009.

The following specification is run separately for each grade and subject:

[7] From now on, "COMPLETE" in capital letters, will refer to the variable, and "complete" in lower case will continue to be used as a noun that describes the type of schedule.
[8] In a relatively few cases, a school reports information for two different schedules, and describes one as complete and the other as a morning or afternoon schedule. This seems unlikely to happen, and it is more likely a mistake that someone makes when filling out the C600 forms. My main sample drops observations for these schools (1,140 observations, or 0.6% of the C600 sample).

$$\text{SABER}_{st} = \beta_0 + \beta_1\text{COMPLETE}_{st} + \beta_2\text{SABER}_{s(t-1)} + \beta_3 X_{st} + d_m + \varepsilon_s \qquad (1)$$

Where $SABER_{st}$ is the average test score in SABER for school *s* in year *t*; $COMPLETE_s$ is a dummy variable equal to 1 if the school has a complete schedule and 0 if it does not (i.e. if it is has a morning or afternoon schedule); $SABER_{s(t-1)}$ is the average test score in the previous period; X_{st} is a vector of school control variables. d_m are municipality fixed effects to control for observed and time-invariant unobserved characteristics that are particular to each municipality.

The main school control variables included are: a dummy for location in a urban area; a dummy for public schools; total school enrollment; teacher-student ratio; the percent of) teachers with a professional education; and the percent of teachers that have specific pedagogic training;[9] and school socioeconomic status (SES), a variable that takes values from 1 to 4, where 1 corresponds to schools with the lowest socioeconomic status, and 4 to the ones with the highest socioeconomic status. The SES is an index that was calculated by the ICFES for each school, using a combination of variables with the proportion of students in the school with each of following characteristics: socioeconomic strata,[10] household resources (e.g., computer, television, and cell phone), household income, dwelling's floor material, parents' education, parents' occupation, number of people in the household, and household overcrowding.

The cross-sectional OLS strategy will yield unbiased estimates of the effect of a complete schedule if the Secretary of Education assigns school schedules and student spots randomly, or based only on observed characteristics. Because the Secretary of Education is supposed to take into consideration a lot of variables when assigning a school schedule and when allocating student spots in each school, the cross-sectional OLS is likely to produce biased estimates of the

[9] The teacher-student ratio, the percent of teachers with a professional education, and the percent of teachers that have specific pedagogic training, are computed for primary (for 5th grade students) and secondary (for 9th grade students).

[10] Colombia classifies all dwellings in six socioeconomic strata according to their geographic location. "Strata 1" corresponds to areas with the lowest socioeconomic characteristics, and "Strata 6" to the areas with the highest socioeconomic characteristics.

impact of COMPLETE. Therefore, I also implement panel OLS and school fixed-effect models to control for additional unobservable characteristics, and mitigate some of the selection issues.

ii. Estimation Strategy 2: Panel with municipality fixed effects

In this model I exploit the panel nature of my data by including indicators for each year to control for any observed or unobserved policy changes or events that affected all schools equally in each year. I also include municipality fixed effects. Because I no longer use lagged tests scores, I am able to include observations from the three years. The model is specified as follows:

$$\mathrm{SABER}_{st} = \beta_0 + \beta_1 \mathrm{COMPLETE}_{st} + \beta_2 X_{st} + d_t + d_m + \varepsilon_{st} \qquad (2)$$

Where $SABER_{st}$ is the average test scores in SABER for school *s* in year *t*; $COMPLETE_{st}$ is a dummy variable equal to 1 if the school has a complete schedule and 0 if it does not (i.e. if it is a half-day schedule); d_t are year fixed effects; d_m are municipality fixed effects; and all other notation represents the same variables as in (1).

iii. Estimation Strategy 3: Panel with school fixed effects

The problem with the cross-sectional OLS and the panel with municipality fixed effects strategies, is that it is possible that the observed differences between the schools with different schedules might simply reflect pre-existing differences between the schools or their students, or reflect the effect of unobserved characteristics that are correlated with having both a complete schedule and higher test scores. Consequently, this strategy may provide a biased estimate of the impact of attending a complete schedule.

In order to mitigate this potential problem, I use the panel dataset to implement a school fixed-effects strategy, by exploiting the fact that some schools switched from having a half-day schedule to a complete schedule, or vice versa. The school fixed effects focuses on analyzing within-school variation in mean school test scores across different cohorts (time), allowing me to break the correlation between unobserved school characteristics and having a complete schedule, and therefore eliminating the bias created by schools systematically selecting into a type of schedule (Bifulco, Fletcher, and Ross, 2011; Hoxby, 2000). To implement this strategy I estimate the following specification:

$$\mathrm{SABER}_{st} = \beta_0 + \beta_1 \mathrm{COMPLETE}_{st} + \beta_2 X_{st} + d_t + d_s + \varepsilon_{st} \qquad (3)$$

The dependent variable is, again, the SABER test score for the school in each period. It is a function of time-varying school characteristics (X), the school schedule in each period (COMPLETE), and time (dt) and school (ds) fixed effects.

The identifying assumption for the school fixed effect model is that schools that switched from having a complete schedule to a single shift, or vice versa, did so for reasons exogenous to the test scores. That is, the reason for the switch must be unrelated to test scores or to unobservable characteristics that are correlated with test scores. I argue that because switching into a different type of schedule is a municipality-level decision typically based on policy or on the demand and supply for school spots, it is exogenous from the point of view of the school, the parents and the students, and unlikely to be correlated with year to year (i.e., across cohort) variations in test scores. In section 6, I use a balancing test to check if the switch to a complete schedule is in fact uncorrelated with school characteristics.

Additionally, given that the Secretary of Education mandates that in the process of allocating spots each year, priority should be given to students that are already enrolled in the schools, parents of students in 5th grade and 9th grade are not likely to request a change to a school with a particular type of schedule because it was switching into a particular schedule, given that they would risk losing their school spot. In impact evaluation terminology, this means that 5th graders and 9th graders did not seek to be treated, or alter their condition (enrollment in their current school) in order to get the treatment (enrollment in a school with a complete schedule).

This is my preferred specification and, in my view, the best estimation of the causal impact of having a complete schedule. By controlling for all observed and unobserved time-invariant school characteristics, this specification decreases the possible bias caused by characteristics that do not change over time within the school, although some endogeneity may remain from time-varying characteristics. For example, there might be some selection issues if the Secretary of Education chooses school schedules based on school attributes, and particularly if these school attributes are related to school performance in the SABER. For example, if very motivated principals or Parents' Associations systematically push and convince the Secretary of Education to implement a complete schedule in their schools, the estimate of the impact of having a complete schedule might be biased, because this type of schools (with more motivated principals or Parents' Associations) are also more likely to have a complete schedule and to perform better in SABER. There might also be selection issues if the Secretary of Education assigns students to schools with a certain type of schedule, based on parents' or students' socioeconomic characteristics, for example, if all relatively poor students are assigned to half-day schools.

There might also be endogeneity from time-varying characteristics if students systematically switch schools after the change in schedule. For example, if more educated parents systematically apply for schools that have switched to a complete schedule, or if they move to an area where most schools have a complete schedule. However, this might not be the case, because in Colombia parents do not have much choice of schools or schedules, and in contrast to the US, they do not tend to move because of the schools in the area. Although this is an empirical question that I am not able to test directly from my data, other authors have cited evidence from the 2007 Quality of Life Survey in Bogota, suggesting that of the 21% of households that move within a 2 year period, just 4% of those cite education or health considerations as the reason for the move (Bonilla-Angel, 2011).

5. Data

To implement the strategies described above, this study exploits an extensive and relatively new panel dataset that contains test score results from SABER 5th and 9th for three periods: 2002, 2005, and 2009.[11] SABER 5th and 9th are nationwide standardized assessments administered every three years to all students in 5th grade and 9th grade in Colombia.[12] According to the Ministry of Education, the main goal of these tests is "to contribute to the improvement of the quality of Colombian education, by carrying out periodic measurements of the development of competencies in students enrolled in basic education, as an indicator of the quality of

[11] More precisely, the SABER 2002 was administered in different moments in 2002-2003, but most tests for 5th grade and 9th grade were conducted in 2002, so I refer to this test as SABER 2002. In the same way, the SABER 2005 was administered in different moments in 2005-2006, but I refer to this test as SABER 2005. All SABER 2009 tests were administered in 2009.

[12] The 2012 application of SABER was extended to evaluate third grade students, but data for this round is not included in this dataset.

education."[13] SABER tests are the best available assessments of primary and secondary student achievement in the country because of their scientific rigor and nationwide coverage. The micro data from these tests has only been made available to researchers in the last few years, making this one of the best and newest available sources to study questions related to student achievement in primary and secondary education in Colombia.

The SABER tests are administered by the Instituto Colombiano para la Evaluación de la Educación (ICFES, the Colombian agency for the evaluation of education) every 3 years. Along with the test, some socioeconomic information of the students and schools is also collected. The SABER 5th and 9th are designed to evaluate competencies in language (Spanish), math, and science[14] at the end of each basic education cycle. Therefore, they are administered to students in 5th grade, which is the end of the "basic primary cycle," and 9th grade, which is the end of the "basic secondary cycle." SABER evaluates student achievement in all schools in the country, both private and public, but private schools are excluded from this analysis because they are not affected my most public policy decisions, including those regarding the length of the school day.

The SABER panel dataset used in this study comes from the ICFES' databases. The SABER 2002 used different evaluation instruments, and the methodology for implementing the tests changed in the three rounds. Therefore, initially it was not possible to find information about performance in SABER that was comparable through time. However, since 2008, the ICFES, together with the World Bank, has worked to reclassify the tests administered in 2002 and 2005, to make them fully comparable to those administered in 2009. As a result of this process, they constructed a dataset with information from student achievement in the 2002, 2005,

[13] Author's translation from the Ministry of Education's website. Retrieved from: http://www.mineducacion.gov.co/1621/w3-article-244735.html, on June 1st, 2013.
[14] The science assessment was introduced in 2012, so it is also excluded from this dataset and the analysis.

and 2009 rounds of SABER (they refer to this dataset as "historic SABER"). This study draws on this dataset because it contains comparable test scores across time, allowing me to compare variations in academic achievement over this period.

The SABER panel contains test scores at the individual level, and some basic identification information, but it does not contain information on school characteristics. Therefore, in order to implement the strategies described above, I merge the SABER panel dataset with school data from the C600, a survey used annually by the Departamento Administrativo Nacional de Estadística (DANE, the national statistics agency) to collect information on basic school characteristics. This dataset contains information on the schools' type (public or private), location (urban or rural), and school characteristics such as enrollment, information on number of transfers, dropouts, grade repetition and promotion in the previous academic year, number of groups or classes by grade, administrative staff, and some information about teachers (e.g., grade they teach, and their educational level). There are two types of C600 forms: C600 A collects the information for each school, and C600 B collects information for each school schedule. That is, if a school has more than one schedule, it fills out one C600 A form, and one C600 B form for each of its schedules or shifts.

The C600 data was initially fragmented in different datasets for each year and for each section or chapter of the C600 form, so I first merge all these parts for each year, and then I create a panel with information from the C600 from 2002, 2005, and 2009.[15] Even though the C600 is collected at the school-schedule level, the data from SABER does not allow identifying the type of schedule that each student attends. Therefore, in order to match these two datasets I have to aggregate the C600 data at the school level. Because schools that do not have a complete,

[15] For administrative and technical reasons, the C600 was not collected in 2003 and 2004.

morning, or afternoon schedule are very uncommon (mostly night and weekend schedules, and about 5.5% of all schedule-level observations), and their student characteristics might be so different than those of students attending traditional "week-day day schools," before aggregating at the school level I drop night and weekend schedule observations from the dataset, and exclude them from the analysis.

The SABER dataset contains information for each student that took the test, but given that the data is not representative at the student level, and that I only have additional data on school level characteristics, I first aggregate the SABER data at the school level. With both the SABER and C600 panels at the school level, I use the school identification codes to merge them and create a school-level panel. SABER test scores cannot be compared across grades, so I create separate samples for 5th and 9th grade tests, including only schools with non-missing math and language scores in the given grade. . Each of these samples contains the school's average SABER math and language test scores for that grade and its corresponding school characteristics from the C600.[16]

To be part of the panel, a school has to have information from both SABER and C600 for a given year, but it does not necessarily have to have information for the three periods. That is, it is an unbalanced panel.[17] For every year, there are a significant number of schools in the C600 that cannot be matched with any school in the SABER dataset. In some cases this is due to problems matching the school identification codes in the two datasets (because of mistakes when typing-in the data, or because of changes in the codes). Additionally, when the ICFES and the World Bank constructed the historic SABER with data from schools in each of the periods, a

[16] Separating schools into four sub-samples by grade and subject (and therefore including schools that are missing one subject test per grade) does not change the results (available on request).
[17] As part of the robustness checks, I also run my models using a balanced panel with schools that appear in the three periods in the database.

large number of school observations were not matched across years, and were not included in the dataset. Therefore, they could not be matched to C600 either.

There are three main reasons why schools could not be matched across time, and are therefore not included in the "historic" SABER dataset that I use. First, although the reclassification exercise made most of the SABER tests comparable across time, a substantial number of the SABER 2002 tests could not be made comparable and were not included in the historic SABER dataset. Second, some school identification codes could not be matched across time because some small schools were merged into larger "education institutions" with a single administration, changing their identification codes. Finally, there was evidence of cheating in some schools, especially in 2002 and 2005, which made impossible to make these schools' test scores comparable across time (ICFES, 2011).

Table 1 contains descriptive statistics of the school panel for the 5th grade and 9th grade samples. In the fifth grade sample, 18.8% of the schools in the sample are located in an urban area. This percentage increases to 25.4% percent in 2005, and goes down to 21.1% in 2009. The percentage of urban schools is much larger for the 9th grade sample, were between 51.1% (2009) and 60% (2005) of schools in the sample are urban. In the 5th grade sample, the average socioeconomic status (between 1.38 in 2002 and 1.41 in 2005) and the high percentage of schools with the lowest socioeconomic status (almost three-quarters in each year) indicates a high concentration of the poorest schools in the school panel. In the 9th grade sample, although there is also a large percentage of schools with the lowest socioeconomic status, the average SES is higher (around 1.8), and the concentration of poorest schools is lower than in the 5th grade sample (less than half the schools in 2002 and 2005, and 52.4% in 2009). In the 5th grade sample, around 27 percent of schools have a complete schedule in 2002 and 2009, and 34.2 percent of schools have a complete schedule in 2005. In the 9th grade sample there is a similar trend: In 2002, 21.5 percent of schools have a complete schedule, and this number goes up to 23.9 percent in 2005, and goes down again in 2009 to 19.9 percent.

Table 2 presents the unconditional mean and standard deviation of test scores, and test scores at different percentiles of the distribution, for schools in the 5th grade and 9th grade samples.[18] In general, the unconditional mean of all test scores goes down in time. The unconditional mean in both grades and both subjects got worse between 2002 and 2005, but the largest decrease was for the 5th grade math test scores, where there was a tenfold decrease in the mean test score (from -0.01, to -0.10). The smallest decrease was in 9th grade math test scores, with test scores going from -0.09 to -0.12 of a standard deviation. The average 9th grade test scores got even worse in 2009: the language test scores for this grade went from -0.22 in 2005, to -0.28 in 2009, and the math test scores went from -0.12 in 2005, to -0.26 in 2009. In contrast, the average language test scores improved in the 5th grade language sample: it went from -0.16 in 2005, to -0.12 in 2009. The standard deviation narrowed down in time in the four samples.

Since my identification strategy relies on schools that switched schedules, Table 3 shows the number of schools that switched by type of change (from half to complete schedule, and from complete to half schedule), and classified by the year in which the switch takes place. In total, 2,262 schools in the 5th grade sample and 464 schools in the 9th grade sample switched schedule at some point between 2002 and 2009, which corresponds to 12 percent and 10 percent

[18] Test scores were standardized by grade, subject, and year, in the complete SABER dataset (i.e., including private schools and schools with schedules different from complete, morning or afternoon, all of which were excluded from the final panel; and also before losing some observations when merging the SABER dataset with the C600 school data). For this reason, the unconditional means in this table are different from zero, and the standard deviations are different from one.

respectively, of the schools that exist in at least 2 periods in the panel.[19] The number of schools that switched scheduled in each sample, corresponds to the schools that are used in the school fixed effect model to identify the effect of COMPLETE.

6. Results

i. Main Results

Table 4 presents the results of the cross-sectional OLS model (1) for the 4 samples. This model exploits the variation of schedule and other variables across schools to estimate how much of the variation between school test scores can be explained by having a complete schedule, controlling for other observable school characteristics, previous performance on the SABER exam, and municipality. For both of the 5th grade samples, the coefficients on COMPLETE are very small and not statistically significant. In contrast, the results for the 9th grade sample show that schools with a complete schedule have language test scores that are 0.126 of a standard deviation higher (column 3), and math test scores that are 0.114 of a standard deviation higher (column 4), than schools without a complete schedule. Most of the school characteristics are also statistically significantly associated with higher test scores, particularly the previous year test scores (associated with about one tenth of a standard deviation higher test scores), the dummy for urban schools, and the school socioeconomic status. The primary teacher student ratio and the percentage of teachers with a pedagogical training are also statistically significant, but only for

[19] There are 237 schools in the 5th grade sample, and 74 schools in the 9th grade sample, that switched from half to complete in 2005 and then back to half in 2009, or that switched from complete to half in 2005, and then back to complete in 2009. These schools are excluded from the calculations in this table, but they are included in the sample used in the regressions

the 5th grade samples. However, the coefficients from these across-school regressions are likely biased because there might be unobserved characteristics that are correlated with test scores and the likelihood of having a complete schedule.

Table 5 presents the main findings for 5th grade test scores by subject, using the panel specifications (2) and (3) described in section 4.[20] The municipality fixed effects specification is reported in columns 1 and 3, and the school fixed effects specification is reported in columns 2 and 4. As mentioned above, the latter is my preferred specification, because most of the characteristics and confounding factors that could bias the estimates of COMPLETE originate from differences between schools with and without a complete schedule. Therefore, by comparing variations in test scores within a school, across cohorts (before and after a scheduling switch), many of the selection and endogeneity concerns are mitigated.[21]

The first row of Table 5 presents the estimates for COMPLETE, the impact of having a complete schedule on test scores. COMPLETE is statistically significant at the 1 percent level for the panel model for 5th grade language, and across the two specifications for 5th grade math. Looking at the language results first, the estimate of COMPLETE in the municipality fixed effects specification shows that having a complete schedule increases language test scores by 0.055 of a standard deviation, compared to schools without a complete schedule in the same municipality and year (column 1). In the school fixed effects specification (column 2), the

[20] I ran regressions of a base model that only includes COMPLETE, and year and municipality fixed effects as explanatory variables. The results for this model tell us the test scores differences between school with a complete schedule and schools with a half schedule: Schools with a complete schedule have language test scores that are on average 0.285 of a standard deviation higher, and math test scores that are 0.274 of a standard deviation higher, than schools with a half schedule.

[21] Because the school fixed effects only captures the changes in test scores that can be explained by time-varying variables, time-invariant variables are excluded from this regression.

estimate of COMPLETE (0.04) is only slightly lower than in the panel specification, but it is no longer statistically significant.

Columns 3 and 4 in Table 5 show the set of findings for 5th grade math. The coefficient for COMPLETE in both the municipality and the school fixed effects specifications is 0.082 and statistically significant, and in both cases, it is larger than the respective specification for language (e.g., 0.082 v. 0.055, using the municipality fixed effects specification). In the school fixed effects model (column 4), the estimate for COMPLETE can be interpreted as the within-school impact of switching to a complete schedule on math test scores. In this case, among schools that switched schedules, the cohorts exposed to a complete schedule have test scores that are 0.082 of a standard deviation higher than cohorts that attended half schedules. The estimated coefficients are very similar across specifications for 5th grade math (0.082), and they do not change much across specifications for the 5th grade language (0.055 in the municipality fixed effects, and 0.044 in the school fixed effects).The fact that the school fixed effects do not substantially alter the estimates, suggest that most of the selection occurs at the municipality level. The coefficients in all of the specifications are larger than those in the cross-sectional OLS model (Table 4, columns 1 and 2). This is consistent with a possible downward bias in the OLS estimates caused by a systematic selection of lower-quality, and therefore lower-demand, schools into a complete schedule, as noted in sections 2 and 3.

Table 6 presents the findings for 9th grade test scores by subject. The first row shows that there is a statistically significant (at the 1% level) difference in test scores between schools that have a complete schedule and schools that have a half schedule. The municipality fixed effects results reveal test score differences of 0.162 for language (column 1) and 0.137 for math (column 3). The school fixed effects results show that among schools that switched schedules, there is also a statistically significant difference in test scores between cohorts exposed to a complete schedule and cohorts that attended half schedules The estimated effect is 0.110 for language (columns 2) and 0.138 for math (column 4), suggesting stronger effects in math, as in the 5th grade sample.

Most notably, all the coefficients for COMPLETE in the 9th grade sample are larger than the respective coefficients for 5th grade, suggesting that having a complete schedule has a larger positive impact on 9th graders than on 5th graders. The 9th grade estimates are more than double the 5th grade estimates in language, and about 50 percent larger in math. The estimates of COMPLETE are also larger than those in the cross-sectional OLS model (Table 4, columns 3 and 4), except in the school fixed specification for language test scores. Again, this might be reflecting a downward bias in the OLS estimates caused by systematic selection of lower-quality schools into complete schedules.

In sum, the most reliable estimates of the impact of a complete schedule, which come from the school fixed effects specifications, show that among schools that switched schedules, the cohorts exposed to complete schedules have higher test scores than cohorts that attended half schedules, and the impact ranges from 0.082 of a standard deviation for 5th grade math test scores, to 0.138 of a standard deviation for 9th grade math test scores.

ii. Balancing Tests

Throughout this paper I argue that the school fixed effect approach provides the most reliable estimates of the impact of attending a school with a complete a schedule. As mentioned in section 4, the identifying assumption for this approach is that switching to a complete schedule must be uncorrelated with the characteristics of the school. To test this assumption, I estimate separate regressions of school characteristics as a function of COMPLETE, cohort (time) fixed effects, and school fixed effects. Table 7 shows the estimated coefficient for COMPLETE for each of these balancing tests. These coefficients tell us if there is a statistically significant difference in each of the school characteristics in column 1, between cohorts exposed to a complete schedule and cohorts exposed to half schedules, in the same school.

As expected, there is a statistically significant difference in school enrollment between schools exposed to complete and half schedules, given that schools with double shifts are likely to have a larger enrollment. There is also a statistically significant difference in the percentage of teachers with a professional degree in the 9^{th} grade sample. The negative coefficient suggests that schools that switch to complete schedules have a lower proportion of professional teachers. This negative selection would likely bias estimates down. Further, the absence of statistically significant estimates for the other schools characteristics suggests that the observable characteristics of the school are not highly correlated with the scheduling switch within schools across cohorts. Although there might still be some correlation between switching to a complete schedule and unobserved characteristics of the school, these results provide supporting evidence for the internal validity of the school fixed effects identification strategy.

iii. Heterogeneous effects and alternate samples

In Table 8, I test for heterogeneous effects by estimating the school fixed effects specification using different subsamples: only rural schools, only urban schools, and only schools classified in each of the four socioeconomic statuses (SES). Columns 1 to 4 present the results for each grade and subject, while each row represents the coefficient of COMPLETE for the specific subsample. Although the smaller samples reduce the power to detect effects, I find a statistically significant impact of COMPLETE for the rural schools in the 5^{th} grade math sample: cohorts exposed to complete schedules in rural schools have test scores that are almost one tenth of a standard deviation higher than cohorts that attended half schedules in rural schools. There is also a marginally statistically significant (10%) impact of being exposed to a complete schedule in schools in urban areas for the 9^{th} grade language sample.

Splitting the sample by socioeconomic status in the lower rows of Table 8, reveals that the impact of having a complete schedule is statistically significant only for the subsample of schools with the lowest socioeconomic status (SES=1), and it is not statistically significant for any of the subsamples containing schools with higher socioeconomic statuses (SES=2, 3, or 4). Among the poorest schools, cohorts exposed to a complete schedule have test scores that are 0.067 of a standard deviation higher in 5^{th} grade language, and 0.093 of a standard deviation higher in 5^{th} grade math, than the cohorts that attended half schedules. There is an even larger impact of attending a complete schedule for the poorest schools in the 9^{th} grade math sample: cohorts exposed to a complete schedule have test scores that are one fifth of a standard deviation higher that cohorts exposed to half schedules. This provides more evidence supporting what has been found in previous studies: having longer school days has heterogeneous impacts on student achievement, and the impact is generally larger for the poorest schools.

Table 9 shows the results of estimating the school fixed effects specification using another set of subsamples. In the first two rows, I run the model using a subsample with only schools that switched from half to complete (first row), and a subsample with only schools that switched from complete to half (second row), in order to analyze if there are symmetric effects of switching from half to complete, and from complete to half. COMPLETE is only statistically significant for the subsample of schools that went from complete to half: for these schools, the cohorts exposed to a complete schedule have 0.214 of a standard deviation higher test scores in 5th grade math, and 0.306 of a standard deviation higher test scores in 9th grade math. The regressions reported in the third row use only the schools that exist throughout the whole period (i.e., using a balanced school panel). In this case, COMPLETE is only statistically significant for the 9th grade language sample (at the 10% level), and the impact is smaller than for the complete sample of schools (0.089 v. 0.11 (Table 6, column 2)).

Finally, the bottom two rows of Table 9 examine "treatment" intensity. I ask whether the impact of attending a complete schedule is different for schools that were exposed to a complete schedule for longer/shorter lengths of time. The fourth row presents the results for regressions that used a subsample of only schools that switched from a half to complete schedule between 2002 and 2005, and kept the complete schedule in 2009. These schools could be considered "early adopters" and were exposed for longer to the "treatment" (i.e., complete schedule). In this case, at least two cohorts within the school were exposed to a complete schedule. It could also mean that students in these schools were exposed to a complete schedule for more than one grade, although it is not possible to know for sure if students changed schools over this period. Despite the very small sample size (444 schools), the results show that cohorts exposed to a complete schedule in schools that were treated for longer have test scores that are 0.274 of a standard deviation higher in 9th grade language, and 0.292 of a standard deviation higher in 9th grade math, than cohorts that attended half schedule.

The bottom row reports the regressions results using a subsample of only schools that switched from half to complete after 2005. That is, it contains schools that had a shorter duration of a complete schedule and a longer duration of a half schedule in the time period studied.

Essentially, students in these schools were "late adopters," and were exposed to the treatment for a shorter period. . There is no statistically significant impact of being exposed to a complete schedule for the schools in this subsample.

7. Conclusions and policy implications

This study provides some of the first evidence that longer school days have an impact on 5th grade and 9th grade academic achievement in Colombia. By using a school fixed effect model that exploits within-school changes in the type of schedule, I am able to control for observed and unobserved time-invariant characteristics of schools, mitigating some of the most critical selection and endogeneity problems that commonly occur when comparing different types of schools. I find that there is a positive impact of having a complete schedule (approximately 2-3 additional hours) on school achievement in 5th grade math, and 9th grade math and language SABER test scores. Results from the school fixed effects model show that among schools that switch schedules between 2002 and 2009, the cohorts exposed to complete schedules have test scores that are about one tenth of a standard deviation higher than cohorts that attended half schedules. This corresponds to approximately a 2.6 percent increase in test scores with respect to the mean for each grade, subject and year.

To put the magnitude of this effect in perspective, the impact of longer school days is smaller than other popular education interventions in Colombia, namely the PACES voucher program, and the contractual schools in Bogota. The PACES voucher program, which has received a lot of attention in the literature because of the use of a lottery to allocate vouchers, had an impact of about 0.2 standard deviations on student test scores relative to students who did not win the lottery (Angrist et al., 2002). While contractual schools in Bogota, which were traditional public schools whose administration was contracted out to reputed, not-for-profit private schools and universities, had an impact of 0.6 and 0.2 standard deviations in math and verbal tests, respectively, relative to traditional public schools (Bonilla-Angel, 2011). However, the magnitude of the impact estimated in this study is similar to other school intervention such as reducing class size by 4 students (Krueger, 1999), or the impact of charter schools on reading test scores (Angrist et al., 2010), and considering that two-tenths of a standard deviation is roughly the score gain associated with one additional school year (Cole et al., 1993), it is a sizable effect. Overall, the results suggest that lengthening the school day may be an effective policy for increasing student achievement.

Further, I find that the impact of having a complete schedule is larger for math test scores than for language test scores in both 5th and 9th grade (e.g., 0.138 v. 0.11 respectively, for 9th graders). This result is not surprising, since, as other research has found, schools may play a larger role in teaching math relative to language, since language skills are often shaped by the home environment.

More notable, is that the positive effects of a longer school day are stronger for 9th grade students than for 5th grade students (e.g., 0.138 v. 0.082 respectively, for math test scores). This makes sense since adolescents may be more likely to engage in risky behaviors outside of school than younger children. Even if they are not engaging in academic endeavors during the extra school hours, 9th graders are certainly spending less time exposed to risk factors outside of school, which ultimately translates into better academic achievement.

Finally, my results suggest that the effects of complete schedules are heterogeneous. Like many other interventions in developing countries, the impact of complete schedules on test scores is largest among the poorest schools and those in rural areas.

Overall, my findings complement those of previous studies in Colombia, which have found a positive impact of attending a complete schedule on dropout and grade repetition in primary (Garcia et al., 2012), and a positive impact of attending a complete schedule on graduation-exit tests (Bonilla-Mejia, 2011). Together, all these findings suggest that longer schools days have a positive impact on academic achievement and other student outcomes in Colombia. Therefore, lengthening the school day may be an effective policy for increasing student achievement, particularly for the lowest-income students in Colombia and other developing countries.

One argument against increasing the length of the school day is that the personnel costs will almost double because of the need to hire new teachers for the additional shift or the additional hours. However, in the Colombian context this might not be the case, because right now different groups of teachers serve different school shifts, so a change to a complete schedule would imply a salary increase for current teachers, but will not require hiring new teachers (Garcia et al., 2012). Similarly, it might be argued that implementing a complete schedule in all public schools would require high investments in infrastructure, and possibly constructing many more schools. However, right now there a large number of schools that have only a morning or an afternoon shift (Bonilla-Mejia, 2011). This means that there is an opportunity for increasing the length of the school day at least in those schools, without high investments in new infrastructure or constructing new schools.

A caveat with the main empirical strategy used in this paper, the school fixed effects model, is that it is not able to control for time-variant characteristics that could be correlated with both the change in schedule and test scores. Although I argue that these are not likely to be large, future research should consider expanding the empirical work to include other quasi-experimental estimation strategies or, if possible, a randomized controlled trial to more definitively assess the impact of longer school days on student achievement.

8. References

Angrist, J., Dynarski, S., Kane, T., Pathak, P., & Walters, C. (2010). Who Benefits from KIPP? NBER, Working Paper 15740.

Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review*, *92*(5), 1535-58.

Barrera-Osorio, F., Maldonado, D., & Rodríguez, C. (2012). Calidad de la Educación Básica y Media en Colombia: Diagnóstico y Propuestas. *Documento CEDE No. 41*.

Berliner, D. (1990). What's all the fuss about instructional time? In Ben-Peretz, M. & Bromme, R. (Eds.), *The nature of time in schools* (3-35). New York: Teachers College Press.

Bellei, C. (2009). Does Lengthening the School Day Increase Students' Academic Achievement? Results from a Natural Experiment in Chile. *Economics of Education Review, 28*(5*),* 29-40.

Bifulco, R., Fletcher, J., & Ross, S. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: economic policy* 3, 25-63.

Bonilla-Angel, J.D. (2011). Contracting out public schools and academic performance. Evidence from Colombia (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.

Bonilla-Mejia, L. (2011). Doble jornada escolar y calidad de la educación in Colombia. Documentos de trabajo sobre Economía Regional, No. 143.

Camacho, A., & Mejía, D. (2013). The Externalities of Conditional Cash Transfers on Crime: The case of Familias en Acción in Bogotá. Universidad de Los Andes, Documento CEDE 2013- 15.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723-733.

Cerdan-Infantes, P., & Vermeersch, C. (2007). More time is better: An evaluation of the full-time school program in Uruguay (Working Paper No. 4167). Retrieved from The World Bank website: http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-4167

Cole, N., Trent, E., & Wadell, D. C. (1993). La prueba de realización en Español: Teacher's guide and technical summary. Chicago: Riverside Publishing Company, 1993.

Cooper, H., Batts Allen, A., Patall, E., & Dent, A. (2010). Effects of full-day kindergarten on academic achievement and social development. *Review of Educational Research, 80* (1), 34-70.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66* (3), 227-268.

Patall, E., Cooper, H., & Batts Allen, A. (2010). Extending the school day or school year: A systematic review of research (1985-2009). *Review of Educational Research, 80* (3), 401-436.

Eren, O., & Millimet, D. (2007). Time to learn? The organizational structure of schools and student achievement. *Empirical Economics, 32*, 301–332.

Gaviria, A., & Barrientos, J. (2001) Determinantes de la calidad de la educación en Colombia. Archivos de Economía, No. 159.

Garcia, S., Fernandez, C., & Weiss, C. (2012). Does lengthening the school day reduce the likelihood of early school dropout and grade repetition: Evidence from Colombia. Paper presented at the Population Association of America 2012 Annual Meeting, San Francisco, CA, May 3-5, 2012.

Hanushek, E., & Wossermann, L. (2007). Education quality and economic growth. Washington, DC: The World Bank.

Hanushek, E. (1979). Conceptual and empirical issues in the estimation of educational production functions. *The Journal of Human Resources*, *14*(3), 351-388.

Harn, B., Linan-Thompson, S., & Roberts, G. (2008). Intensifying instruction: Does additional time make a difference for the most at-risk first graders? *Journal of Learning disabilities*, *41*(2), 115-125.

Hoxby, C. (2000b). Peer effects in the classroom: Learning from gender and race variation (Working Paper No. 7867). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w7867

ICFES (2011). SABER 5º y 9º 2009. Informe técnico de resultados históricos. Bogotá, D.C.

Jacob, B., & Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. *American Economic Review*, 93, pp. 1560-1577.

Krueger, A. (1999). Experimental estimates of education production function. *The Quarterly Journal of Economics*, *114*(2), pp. 497-532.

Kruger, D. I., & Berthelon, M. (2009). Delaying the bell: The effects of longer school days on adolescent motherhood in Chile. *Institute for the Study of Labor Discussion Paper, 4553*.

Lavy, V. (2010). Do differences on school's instruction time explain international achievement gaps in Math, Science and Reading? Evidence from developed and developing countries (Working paper series No. 16227). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w16227

Lee, J. & Barro, R. J. (2001), Schooling Quality in a Cross–Section of Countries. *Economica, 68*, 465–488

Llach, J., Adrogue, C., & Gigaglia, M. (2009). Do Longer School Days Have Enduring Educational, Occupational, or Income Effects? A Natural Experiment in Buenos Aires, Argentina. *Economía, 10* (1), 1-43.

Marcotte, D. (2007). Schooling and Test Scores: A Mother-Natural Experiment. *Economics of Education Review, 26*(3), 629-40.

Marcotte, D., & Hansen, B. (2010). Time for School. *Education Next, 10* (1), 52-59.

Margo (1994). Explaining Black-White Wage Convergence, 1940-1950. *Industrial and Labor Relations Review, 48* (3) (Apr.1995), 470-481

McMullen, S., & Rouse, K. (2012). The impact of year-round schooling on academic achievement: Evidence from mandatory school calendar conversions. *American Economic Journal: Economic Policy, 4* (4), 230-252.

Ministerio de Educación Nacional (2013). Pruebas SABER. Retrieved from: http://www.mineducacion.gov.co/1621/w3-article-244735.html, on June 1st, 2013.

Ministerio de Educación Nacional (2006). Cartilla 1: Proceso de matrícula. Retrieved from: http://www.mineducacion.gov.co/cvn/1665/article-99324.html on November 14th, 2013.

Olsen, D., & Zigler, E. (1989). An assessment of the all-day kindergarten movement. *Early Childhood Research Quarterly, 4*, 167–186.

Pires, T., & Urzua, S. (2011). Longer schools days, better outcomes? (Working Paper). Northwestern University.

Rathburn, A. (2010). Making the most of extra time: Relationship between full-day kindergarten instructional environments and reading achievement. American Institute for Research, Research Brief.

Redd, Z., Boccanfuso, C., Walker, K., Princiotta, D., Knewstub, D., & Moore, K. (2012). Expanding time for learning both inside and outside of the classroom: A review of the evidence base (Child Trends #2012-20).

Rivkin, S., & Schiman, J. (2013). Instruction time, classroom quality, and academic Achievement. NBER Working Paper No. 19464.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 4147-58.

Valenzuela, J.P. (2005). Essays in economics of education. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.

World Bank (2008). Quality of education in Colombia. An analysis and options for a policy agenda. Report No. 43906-CO.

Table 1. Descriptive statistics of school panel

	5th grade			9th grade		
	2002	2005	2009	2002	2005	2009
Total schools	14419	15006	24767	2970	4214	6428
% urban	18.8%	25.4%	21.1%	56.4%	60.0%	51.1%
Average socioeconomic status (SES)	1.38	1.41	1.39	1.80	1.89	1.77
% of poorest schools (SES=1)	73.5%	72.8%	72.9%	49.7%	46.0%	52.4%
% of richest schools (SES=4)	2.2%	2.8%	2.4%	7.4%	8.2%	7.0%
Average school enrollment	195	258	205	609	766	673
Average 5th/9th grade enrollment	22	30	24	62	82	73
Average primary/secondary enrollment	129	153	118	321	387	342
Primary/secondary teacher-student ratio	0.04	0.04	0.05	0.07	0.06	0.06
% primary/secondary teachers with professional degree	59.4%	68.3%	67.0%	86.5%	91.7%	90.0%
% primary/secondary teachers with pedagogic training	60.5%	67.8%	67.0%	83.8%	86.9%	83.5%
% schools with complete schedule	27.0%	34.2%	27.1%	21.5%	23.9%	19.9%
% schools that switched schedule	14.1%	10.4%	9.8%	15.3%	11.4%	8.2%

Note: School socioeconomic status (SES) is an index that goes from 1 (poorest schools) to 4 (richest schools).

Table 2. Distribution of test scores by year

	Language			Mathematics		
	2002	2005	2009	2002	2005	2009
Percentile			**5th grade**			
5%	-1.53	-1.63	-1.49	-1.48	-1.53	-1.45
25%	-0.76	-0.78	-0.74	-0.84	-0.79	-0.77
Median	-0.14	-0.21	-0.21	-0.10	-0.24	-0.23
75%	0.64	0.38	0.39	0.73	0.46	0.40
95%	1.75	1.61	1.64	1.77	1.80	1.71
Mean	**-0.04**	**-0.16**	**-0.12**	**-0.01**	**-0.10**	**-0.11**
St. deviation	**1.01**	**0.96**	**0.94**	**1.02**	**1.00**	**0.95**
Percentile			**9th grade**			
5%	-1.37	-1.63	-1.59	-1.12	-1.06	-1.43
25%	-0.73	-0.75	-0.78	-0.73	-0.61	-0.77
Median	-0.26	-0.24	-0.24	-0.38	-0.34	-0.29
75%	0.30	0.28	0.21	0.19	0.06	0.16
95%	1.73	1.28	0.95	2.16	1.78	1.08
Mean	**-0.13**	**-0.22**	**-0.28**	**-0.09**	**-0.12**	**-0.26**
St. deviation	**0.94**	**0.89**	**0.80**	**0.99**	**0.94**	**0.80**

Note: Test scores are standardized by grade, subject and year.

Table 3. Number of schools that switched schedule by type of change and year

	5th grade	9th grade
Total number of schools*	19546	4570
Number of schools that switched schedule	**2,262**	**464**
Percent of schools that switched schedule	12%	10%
From half to complete schedule	**1398**	**250**
half in 2002, complete in 2005 and 2009	409	148
half in 2002 and 2005, complete 2009	120	50
half in 2002, complete in 2005**	47	8
half in 2002, complete in 2009**	723	25
half in 2005, complete in 2009**	99	19
From complete to half schedule	**864**	**214**
complete in 2002, half in 2005 and 2009	173	75
complete in 2002 and 2005, half in 2009	89	38
complete in 2002, half in 2005**	14	3
complete in 2002, half in 2009**	220	33
complete in 2005, half in 2009**	368	65

Notes: Schools that switched from half to complete in 2005 and then back to half in 2009, or that switched from complete to half in 2005, and then back to complete in 2009 are excluded from the total number of schools that switched schedule (237 schools in the 5th grade sample, and 74 schools in the 9th grade sample).*Total number of schools that exists in 2 or 3 periods. **These schools only exist in 2 periods.

Table 4. Cross-sectional OLS regressions. Dependent variable: School test scores

	5th grade		9th grade	
	Language	Mathematics	Language	Mathematics
	(1)	(2)	(3)	(4)
Complete	0.006	0.040	0.126***	0.114**
	[0.033]	[0.036]	[0.041]	[0.047]
Test score in previous period	0.087***	0.108***	0.115***	0.084***
	[0.008]	[0.008]	[0.015]	[0.018]
Dummy for urban schools	-0.097***	-0.127***	-0.049	-0.031
	[0.028]	[0.027]	[0.031]	[0.032]
School socioeconomic status	0.137***	0.111***	0.297***	0.205***
	[0.020]	[0.018]	[0.018]	[0.020]
School enrollment	0.001	-0.005**	0.001	-0.001
	[0.002]	[0.002]	[0.002]	[0.002]
Primary teacher-student ratio	1.270***	0.490	0.093	0.323
	[0.346]	[0.348]	[0.328]	[0.312]
% teachers with professional education in primary	0.048	0.049	0.006	-0.087
	[0.031]	[0.032]	[0.071]	[0.075]
% teachers with pedagogical training in primary	0.064**	0.062**	0.024	-0.061
	[0.031]	[0.031]	[0.065]	[0.076]
Observations	27,578	27,578	6,984	6,984
R-squared	0.209	0.228	0.440	0.312
Number of municipalities	1035	1035	1012	1012

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Notes: All regressions include municipal fixed effects. Complete is a dummy variable equal to 1 when the school has a complete schedule. School socioeconomic status (SES) is an index that goes from 1 (poorest schools) to 4 (richest schools).

Table 5. Panel regression results. Dependent variable: 5th grade test scores.

	Language		Mathematics	
	Municipality fixed effects	School fixed effects	Municipality fixed effects	School fixed effects
	(1)	(2)	(3)	(4)
Complete	0.055**	0.044	0.082***	0.082**
	[0.024]	[0.034]	[0.027]	[0.033]
Dummy for urban schools	-0.101***	--	-0.181***	--
	[0.022]		[0.024]	
School socioeconomic status	0.115***	--	0.078***	--
	[0.016]		[0.014]	
School enrollment	-0.002	0.007	-0.009***	0.004
	[0.001]	[0.006]	[0.003]	[0.006]
Primary teacher-student ratio	1.271***	0.509	0.791***	-0.123
	[0.277]	[0.468]	[0.264]	[0.470]
% teachers with professional education in primary	0.033	-0.036	0.039	-0.016
	[0.026]	[0.047]	[0.025]	[0.046]
% teachers with pedagogical training in primary	0.067***	0.048	0.052**	0.041
	[0.026]	[0.046]	[0.024]	[0.045]
Observations	54,192	54,192	54,192	54,192
R-squared	0.161	0.611	0.185	0.618
Year fixed effects	YES	YES	YES	YES
Municipal fixed effects	1093	NO	1093	NO
School fixed effects	NO	26919	NO	26919

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Notes: Complete is a dummy variable equal to 1 when the school has a complete schedule. School socioeconomic status (SES) is an index that goes from 1 (poorest schools) to 4 (richest schools). All standard errors are robust.

Table 6. Panel regression results. Dependent variable: 9th grade test scores.

	Language		Mathematics	
	Municipality fixed effects	School fixed effects	Municipality fixed effects	School fixed effects
	(1)	(2)	(3)	(4)
Complete	0.162***	0.110**	0.137***	0.138**
	[0.033]	[0.055]	[0.034]	[0.063]
Dummy for urban schools	-0.040	--	-0.074***	--
	[0.029]		[0.025]	
School socioeconomic status	0.269***	--	0.168***	--
	[0.021]		[0.017]	
School enrollment	0.001	-0.016***	-0.001	-0.008
	[0.002]	[0.006]	[0.002]	[0.007]
Secondary teacher-student ratio	0.230	0.032	0.548*	0.451
	[0.296]	[0.573]	[0.319]	[0.503]
% teachers with professional education in secondary	0.036	-0.028	-0.014	-0.076
	[0.052]	[0.091]	[0.052]	[0.097]
% teachers with pedagogical training in secondary	-0.005	-0.103	-0.034	-0.056
	[0.047]	[0.090]	[0.051]	[0.097]
Observations	13,612	13,612	13,612	13,612
R-squared	0.287	0.661	0.212	0.611
Year fixed effects	YES	YES	YES	YES
Municipal fixed effects	1,083	NO	1,083	NO
School fixed effects	NO	6725	NO	6725

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Notes: Complete is a dummy variable equal to 1 when the school has a complete schedule. School socioeconomic status (SES) is an index that goes from 1 (poorest schools) to 4 (richest schools). All standard errors are robust.

Table 7. Balancing tests for COMPLETE

	5th grade	9th grade
	(1)	(3)
Enrollment	-0.143***	-0.454***
	[0.018]	[0.091]
Teacher student ratio	0.001	-0.000
	[0.001]	[0.003]
% teachers with professional degree	0.001	-0.025**
	[0.010]	[0.011]
% teachers with pedagogic training	-0.005	-0.014
	[0.010]	[0.013]

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1
Note: All regressions include school and year fixed effects. Each column reports the coefficient for COMPLETE, which shows the differences in school characteristics (the dependent variable in the first column) between complete and half schedule schools, controlling for school and year.

Table 8. Heterogeneous effects: School fixed effects model using alternate samples

	5th grade		9th grade	
	Language	Mathematics	Language	Mathematics
	(1)	(2)	(3)	(4)
Rural schools	0.056	0.097***	0.116	0.175*
	[0.037]	[0.036]	[0.092]	[0.103]
Observations	*42,446*	*42,446*	*6,121*	*6,121*
Urban schools	0.005	0.009	0.104*	0.106
	[0.068]	[0.077]	[0.063]	[0.073]
Observations	*11,746*	*11,746*	*7,491*	*7,491*
Poorest schools (SES=1)	0.067*	0.093**	0.138	0.200**
	[0.039]	[0.038]	[0.084]	[0.093]
Observations	39,571	39,571	6,785	6,785
Middle-to-poor (SES=2)	0.006	0.087	0.103	0.102
	[0.077]	[0.075]	[0.081]	[0.094]
Observations	*9,298*	*9,298*	*3,564*	*6,785*
Middle-to-rich (SES=3)	0.072	-0.028	0.027	0.001
	[0.134]	[0.143]	[0.152]	[0.162]
Observations	*4,012*	*4,012*	*2,251*	*3,564*
Richest schools (SES=4)	-0.169	-0.132	0.121	-0.029
	[0.142]	[0.177]	[0.170]	[0.330]
Observations	*1,311*	*1,311*	*1,012*	*2,251*

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Notes: Complete is a dummy variable equal to 1 when the school has a complete schedule. School socioeconomic status (SES) is an index that goes from 1 (poorest schools) to 4 (richest schools). Only the coefficient for COMPLETE is reported. All regressions control for school enrollment, teacher student ratio, percentage of teachers with a professional degree, percentage of teachers with pedagogic training, and include school and year fixed effects. All standard errors are robust.

Table 9. School fixed effects model using alternate samples

	5th grade		9th grade	
	Language	Mathematics	Language	Mathematics
	(1)	(2)	(3)	(4)
Switchers from half to complete	0.085	0.055	0.017	-0.263
	[0.114]	[0.113]	[0.155]	[0.180]
Observations	*3,325*	*3,325*	*698*	*698*
Switchers from complete to half	0.054	0.214*	0.079	0.306*
	[0.128]	[0.126]	[0.167]	[0.179]
Observations	*1,990*	*1,990*	*541*	*541*
Schools in 3 periods (balanced panel)	-0.024	-0.017	0.089*	0.096
	[0.041]	[0.039]	[0.053]	[0.063]
Observations	*23,181*	*23,181*	*6,951*	*6,951*
Switchers with longer treatment	0.101	-0.055	0.274***	0.292***
	[0.076]	[0.077]	[0.103]	[0.105]
Observations	*1,227*	*1,227*	*444*	*444*
Switchers with shorter treatment	0.020	0.134	-0.036	-0.030
	[0.121]	[0.134]	[0.207]	[0.182]
Observations	360	360	150	150

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Notes: Complete is a dummy variable equal to 1 when the school has a complete schedule. School socioeconomic status (SES) is an index that goes from 1 (poorest schools) to 4 (richest schools). Only the coefficient for COMPLETE is reported. All regressions control for school enrollment, teacher student ratio, percentage of teachers with a professional degree, percentage of teachers with pedagogic training, and include school and year fixed effects. Switchers with longer treatment are schools that were half in 2002, switched to complete in 2005 and kept being complete in 2009. Switchers with shorter treatment are schools that were half in 2002 and 2005, and switched to complete only in 2009. All standard errors are robust.