Do longer school days improve student achievement? Evidence from Colombia.

June, 2013

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 or a half-day schedule. Using test score data from 5th and 9th graders in 2002-2003, 2005-2006, and 2009, along with school administrative data, this study identifies the effect of longer school days by implementing a school fixed effects model. The base model compares changes in average test scores for schools that switched from a complete schedule to a half schedule and vice versa. I also use the demand and supply for school spots in a municipality to instrument the probability of having a complete schedule. I find that schools with a complete schedule have about one tenth of a standard deviation higher test scores than those with half schedules. The impact 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. These results suggest that policies that aim to increase the length of the school day could improve student achievement in developing countries.

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

JEL codes: I21, I28, H43.

I. Introduction

There are large differences in the time students spend in school, both across and within countries. Are these differences contributing to the differences in student learning and achievement gaps? 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 adding more hours of school per day or more days of school per year will improve learning outcomes.

Recently, some studies have addressed this gap in the literature and analyzed the effect of more instructional time on student achievement, and they have found some evidence that extending the time spent in school might raise student achievement. Most of these studies have focused on the length of the school year, finding in general positive, small and heterogeneous impacts (e.g. Marcotte, 2006; Eren and Millimet, 2007; Hoxby and Murarka, 2009).

In Latin America, where some countries like Chile and Argentina have experimented with increasing the length of the school day, the literature has focused on analyzing the impact of longer school days. The results of these studies suggest that attending longer school days has a positive impact on student achievement. However, the impact is different for different types of students and schools. These differences depend mostly on the content of the additional hours (Llach, Drogue, and Gigaglia, 2009) and on the students' and schools' socioeconomic characteristics (Valenzuela, 2005). This literature will be discussed in more detail in the next section.

In Colombia there are differences in the length of school days for students attending primary and secondary grades: some students attend school for a full day or complete schedule (what is known in the country as "jornada única" or "jornada completa"), while some others attend school only for half a day or schedule, either in a morning or afternoon shift.[1] It is commonly believed that extending the time that students spend in school will improve their academic achievement, and could also benefit students by keeping them "out of trouble," reducing crime rates and teen pregnancies. Part of this belief comes from the fact that students who attend schools that have a complete schedule sometimes appear to perform better than those who attend schools that have a morning or afternoon shift, even though this result can be due to other characteristics associated with students attending that particular school (e.g. higher income, more educated parents, etc.).

Arguing that extending the school schedule might benefit students, policymakers have proposed, on several occasions, the implementation of a full school day or complete schedule in all public schools, which have in many cases double-shift schedules.[2] For example, in 1994, the government passed the General Education Law (Law 115), which established that all public schools should have a single complete or full-day schedule. However, plans to implement this law did not begin until years later and were completely abandoned by 2002. This happened partly because of a substantial increase in demand for public schools that was difficult to meet because of low installed physical capacity (fewer schools than the ones needed to meet the

[1] In some schools, primary students attend the morning shift and secondary students attend the afternoon shift or vice versa. In other cases, schools have only one shift, with either primary or secondary students attending that shift.
[2] Two different groups of students attend the same school, with one group attending the morning shift, and the other one the afternoon shift.

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 public schools in at least two of the largest cities in the country, Bogota and Cali, and pilot programs are currently being implemented. In April 2012, Bogota city announced a pilot to turn 25 public schools into complete schedule schools, and Cali's Mayor announced that 2 schools in the city (approximately 1,000 students) will start having a "complementary schedule" and that this new standard will be implemented in the entire city during his term. The complete or complementary schedule in these two cities will mainly consist of three extra hours in school devoted to civics, sports, and cultural activities, and not to additional instructional time. However, the argument for implementing this increase in the length of the school day is that more time spent in school will translate into better outcomes for students. There are mainly three different arguments of why longer schools days could benefit students: First, they will spend less time alone at home doing educationally unproductive activities. Second, they will spend less time in the streets, exposing themselves to different risks. And third, by spending more time in school they will get devote more time to learning.

If longer school days have the capacity to benefit students, families, and even the country as whole, and these benefits are larger than the costs, then it makes sense for a government to try to extend the school day for all students. However, it is very difficult to measure these benefits. Generally, the belief that students attending longer schools days do better is based on observed differences in student achievement between schools with different schedules, but these differences can reflect differences in many other characteristics, and not only in the length of the school day or in instructional time. For example, most private schools have a complete schedule, so the observed differences in student outcomes between schools with a complete schedule and schools with a single shift, might be reflecting other characteristics that are systematically present in private schools, such as higher income and more educated parents, and better health. Even among public schools, it is possible that parents with particular characteristics might systematically enroll their children in schools that have a complete schedule, or schools with complete schedules might be more likely to be present in particular areas or municipalities. I address these issues in detail in section III, but generally, I find little support for selection of parents into a particular schedule in public schools.

In this context, this paper analyzes the impact of longer school days on student achievement in Colombia by exploiting the differences in the length of the school day for 5th grade and 9th grade students attending a school in the country. I find that schools with a complete schedule have about one tenth of a standard deviation higher test scores than those with half schedules. The impact 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.

To my knowledge, this is one of the first studies to look at the impact of longer school days on student achievement, and the first to look at the impact on student achievement in primary and secondary students in Colombia.[3] The main contribution of this study is to estimate the impact of longer school days using a variety of econometric approaches to address possible selection or endogeneity problems. Additionally, given that most of the literature on the impact of instructional time on student learning has focused on analyzing the impact of differences in the length of school year, this study will contribute to the literature by analyzing the differences in instructional time given by the length of the school day instead. Furthermore, it will also

[3] In Colombia, the education system is divided in three levels: preschool (at least one year); basic education, which includes basic primary (1st -5th grade) and basic secondary (6th- 9th grade); and middle education or middle secondary (10th- 11th grade).

contribute to the literature on the effect of instructional time on student achievement in a developing country, which is very scarce, and in Colombia, where it is practically nonexistent. Given the particular situation in which, for example, a 5th grader could be attending school for 7 hours while another similar 5th grader could be attending school for 4 hours for reasons arguably exogenous to the family's choice, I believe that this provides a reasonable quasi-experiment of the impact of longer schools days on student achievement.

The rest of the paper is organized as follows: Section II presents a brief conceptual framework and reviews the most relevant literature on instructional time. Section III discusses the identification strategies. Section IV describes the data. Section V presents results. And Section VI concludes.

II. Conceptual framework and literature review

The literature that has analyzed the impact of instructional time goes back to the 1960s, when Carroll (1963, as cited in Llach et al., 2009) analyzed variations in school learning as a function of five classes of variables, three of them expressed in terms of time (aptitude; opportunity to learn –amount of time allowed for learning, for example, by school schedule or program; and perseverance) and two in terms of achievement (quality of instruction and ability to understand instruction). Carroll concluded that the learner would be successful in learning something to the extent that he spends the amount of time that he needs to learn the task, so increasing instructional time has a positive but small impact on educational achievement. The impact tended to be higher the lower the country's GDP and the students' socioeconomic status.

Cotton (1989, as cited in Llach, et al., 2009) did a comprehensive review of this literature, and she identified different kinds of allocated time: school time (time spent in school), classroom time (time spent in the classroom), instructional time (time spent teaching), engaged time (time when students pay attention and attempting to learn), and academic learning time (portion of engaged time that students spend working on tasks at an appropriate level of difficulty for them and experiencing high levels of success). She concluded that there is a small, but positive relationship between allocated time on task and achievement, and a strong and positive association between academic learning time and student achievement and attitudes. She also found that students with lower ability benefit more from increases in instruction time.

Most of this pioneering literature analyzed the impact of school time mainly from the point of view of educational psychologists or educators, focusing on analyzing the learning process inside the class. This literature was mainly a descriptive and focused on developed countries, and relatively little was known about the impact of instructional time in developing countries.

More recent 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 improves math and science 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 US. 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. In this study, I will also explore the issue of heterogeneous impacts of longer school days.

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 US 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 just under 0.2 standard deviations.

Another part of the literature has focused on analyzing the impact of longer school days or more instructional time per day. For example, in a cross-country study Lavy (2010) found evidence that instructional time consistently had a positive and significant effect on PISA 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. In the U.S., the literature on the length of the school day has focused on studying 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 had a positive association with academic achievement, compared to a half-day kindergarten, but the effect generally seemed to fade-out by third grade. They also found a substantial association 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 such a positive 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.

Some authors suggest that interventions that aim to increase instructional time early in school might have positive effects particularly for the most at-risk or poor children. For example, Olsen and Zigler (1989, as cited in Cooper et al., 2010) found that in general, extended-day programs in kindergarten did seem to increase standardized test scores in the short-term, particularly for disadvantaged, bilingual, or "least-ready" for school children. However, they concluded that there was little evidence that positive changes in academic performance were maintained beyond the early elementary school years or that there were changes in motivation or general intellectual ability that were likely to support significant long-term changes. 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 urban setting than those in non-urban communities. They suggested that children attending 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 (from 30 to 60 minutes) on early literacy skills for the most at-risk first graders. The authors concluded that the students receiving more intensive interventions (more instructional time) made significantly more progress across a range of early reading measures.

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 little variation in the length of the school day or the school year across schools districts (in the U.S.), or municipalities (in other countries), or to variations in length that are directly related to resources, which make difficult to estimate the impact of longer school days or longer school years. By exploiting the differences in the length of the school day among schools in Colombia, this study contributes to the understanding of the impact of longer schools days on primary and secondary student outcomes.

In Latin America, a few studies have exploited natural experiments given by the implementation of a full-day schedule in schools that previously had double-shifts or just morning or afternoon schedules. Valenzuela (2005) did one of the first impact evaluations of the Chilean full school day program. 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 Differences-in-Differences model at the school and at the county level. 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 school day program, Bellei (2009) found that it had positive effect on students' academic achievement in both mathematics and language, and the program 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.

In Argentina, Llach et al. (2009) analyzed the long-term impact of attending longer days in Buenos Aires. They found that students that attended full-day primary schools had a secondary school graduation rate 21 percent higher than those that attended single-shift primary schools, and this was mostly driven by students coming from low socio economic backgrounds. However, they did not find enduring effects on income and employment. Their study compared the content of the classes in single-shift and full-day schools, allowing them to conclude that the content of the additional hours was even more important than increasing the duration of the school day.

To the best of my knowledge, there are only two studies that have evaluated the impact of instructional time in Colombia. In the first one, Bonilla-Mejia (2011) evaluated the impact of attending a complete schedule school on student performance on a high school exit exam (SABER 11, previously known as "Icfes"), using an instrumental variables strategy. 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 morning or afternoon shifts. Additionally, they found that the impact was larger when compared with students attending only an afternoon shift. In the second study, Garcia, Fernandez and

Weiss (2012) used family fixed effects to analyze the impact of changing from a half-day to a complete schedule on early dropout and grade repetition. They found that a complete schedule reduces the probability of early dropout and grade repetition. This study provides the first estimates of the impact of longer school days on primary or secondary student achievement.

The extant literature shows that in general, there are some positive effects of attending school for a longer school year or for longer school days. The evidence suggests that the impact on student outcomes is heterogeneous, and it tends to be more significant for students from poorer backgrounds. There are two main reasons why the impact of time could be different for students from different socioeconomic backgrounds. The first is that the quality of the additional time in school matters. As Llach et al. (2009) show for Buenos Aires, the content of the additional hours is probably more important that 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. The net impact of additional time will also depend to some extent on the quality of teachers. If less prepared teachers are brought to schools in order to be able to extend the number of school hours, then the additional hours spent in school will not necessarily translate into more learning for the students. If this is the case, then those students who are attending longer school days will not necessarily benefit from those additional hours.

The second reason why additional hours in school could have a heterogeneous impact on student outcomes, is that the net impact will depend on the counterfactual of what kids would have been doing while out of school, had they not been in school for those additional hours. One hypothesis is that this counterfactual is likely to vary a lot depending on the socioeconomic characteristics of the student. If not in school for those additional hours, students from higher-income families might be at home, playing, and spending time with their parents doing homework, reading books, etc. In contrast, students from lower-income families might spend more time outside of school and of their homes, increasing their exposure to risk factors associated with crime, teen pregnancies or drugs. Therefore, if the latter spend the additional hours in school in a productive way, they are more likely to benefit from those additional hours that the former.

Given this framework, the impact of longer school days on student achievement in Colombia is uncertain. On one hand, it can 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 on higher student achievement. Therefore, I expect to find a positive impact of longer schools days on student achievement.

On the other hand, the additional hours spent in school could have no impact at all or even a negative impact on student achievement, if the additional hours in school are of poor quality and/or the content is largely irrelevant. For example, if schools that have a complete schedule in Colombia devote the additional hours to sports, or more playing time, then there might not be a significant difference in student achievement between schools that have a complete schedule and schools that only have a morning or afternoon shift. Also, the additional hours could have no impact if those attending the morning or afternoon shifts systematically come from better-off families or communities and they can make up for the fewer hours spent in school.

Additionally, I expect to find heterogeneous impacts for different groups of students or schools, but the direction of these impacts is unclear. If the quality of the additional hours is better for those students attending relatively better schools, the impact of longer school days will have a larger impact on the better-off schools with relatively more advantaged students. However, if the content of the additional hours is designed in a way that benefits mostly disadvantaged students, then the additional hours will have a larger impact on these students' achievement. For example, if the additional hours are assigned to strengthening math and reading skills, the disadvantaged students will have an opportunity to "catch up" with their more advantaged classmates, and the impact of the additionally hours will be larger for them. Regardless of the impact on student achievement, low-income students might benefit more from spending more time in school, just because that implies spending less time in the streets, or in unproductive activities. However, I cannot test directly for these other possible impacts; I focus on only one outcome: student achievement on the SABER tests.

III. Empirical strategy

The main goal of this study is to analyze the impact of longer school days (i.e. complete schedule) on student achievement. If students were randomly assigned to a complete schedule or to half schedules (morning or afternoon sessions), in principle one could estimate the impact of attending a complete schedule by comparing the average achievement of students attending complete schedules with the average achievement of those attending a morning or afternoon session. However, the process of "allocating" students to different schedules and the decision about the type of schedule a school has are not simple, and definitely not completely random.

This generates concerns about endogeneity in the impact estimations. Below, I explain some of the possible endogeneity problems.

First of all, there might be some cases in which all or most of the schools in the municipality have the same type of schedule. This could happen if decisions are taken from a higher level of government, like the Secretary of Education of the department, and based on certain characteristics, such as the need or the resources available in each municipality (A department is similar to a state in the U.S., and is divided in municipalities. There are 32 departments, one capital district (Bogota), and 1,119 municipalities in Colombia). It could also happen if the decisions are taken at the municipal level, and the municipality decides that all its schools will have the same schedule because of its particular situation. For example, it lacks the resources to hire more teachers, and therefore implements single-shifts in all its schools. I will further investigate allocation and variation at the municipal level in the paper.

Second, when there is variation on the type of school schedules within a municipality, the Secretary of Education is usually the one that allocates students to a particular schedule. On one side of this process, when parents want to enroll their children for the first time into the system, or want to transfer their children into a different school because they moved to a new area, they need to fill in an application with the municipality's Secretary of Education. In this application, parents are able to explicitly express their preference for two schools, but they cannot express their preference for a particular schedule.

On the other side, the Secretary of Education makes a decision on where to assign the new or transferred students whose parents have applied for a school spot. Ideally, this decision matches the parents' expressed preferences. However, in practice, it is a decision that has to do mostly with the projections of availability of school spots in the municipality's schools (Ministerio de Educacion Nacional, 2006, Cartilla 1: Proceso de matricula). So 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. If that is the case, the municipality will try to allocate the students to the school closest to their place of residence. However, there is some possibility 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.

At the same time, the supply of school spots for each year is mainly a result of the capacity that schools have. Every year, the Secretary of Education assesses the adjustments in the number of school spots necessary to keep enrolled students in the educational system (and ideally in the same schools they are already attending), and to receive new students who are expected to join the public education system in the following year. In other words, the planning of the school supply depends on the behavior of the demand for schooling from the population in school age. Based on this demand (from the census or surveys), the Secretary of Education determines where should there be an increase in the supply i.e. the number of schools spots, or which schools should offer a complete or half day schedule. As a result, the process of "allocating" students to complete schedule or morning/afternoon sessions is a function of many variables. For example, the number of people who are at school age in the municipality, the actual demand for school spots (i.e. the number of applications filled out by parents), 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).

One could also argue that parents could self- select into a particular schedule, if for example, when they apply for a school spot, they express their preferences for two schools with the same schedule. For example, more educated parents could systematically select as their two preferred options schools that have a complete schedule. There could also be some selection problems if the Secretary of Education wants to, and is able to, assign school spots according to parents' or students' socioeconomic characteristics. For example, if all relatively poor parents are assigned to double-shift schools.

On the other hand, I argue that there are some reasons that mitigate some of the endogeneity concerns described above. First, parents cannot request to be located or apply for a specific school schedule, or even express their preferences in terms of the type of schedule. Second, even though they are able to express their preferences for two schools and therefore could indirectly express their preferences for a specific schedule by applying to schools with the same schedule, in practice they do not have much choice of schools or schedules. Even though parents might derive utility from other things, like school quality and safety, parents' applying for a public school spot are usually low-income parents, so their main decision is about getting their children to enroll in a school, while minimizing the cost of doing so. If we assume that parents want to minimize the costs of enrolling their children in school, then they will try to minimize transportation costs to the school, and they will be more likely to apply to schools relatively close to their place of residence, regardless of their schedule. If this assumption is true, then in practice they do not have many schools with complete schedule and many schools with morning or afternoon shifts from which to choose to apply. Arguably, parents could also move to another area in order to attend a particular school, as it usually happens in the U.S. However, this does not seem to be the case in Colombia. For example, according to the Quality of Life Survey, few people actually cite schools as a reason for moving (J.D. Bonilla, 2012).

Third, the final decision about where children will attend is out of the parents' control. It is the Secretary of Education of the municipality who makes the decision about where each student should attend and which schools will have a complete schedule or a morning and/or afternoon shift, and this decision depends on several variables related mainly to the demand and supply of schools spots. On one hand, each school tells the Secretary of Education of the municipality the number of spots it is able to provide. On the other hand, the Secretary looks at the expected demand for school spots. Finally, based on these estimates of the demand and supply and the resources available in the municipality, the Secretary might decide to divide a complete schedule school into a morning and afternoon shifts, or to merge two shifts into one complete schedule.

Finally, in the school fixed effects model (described in subsection ii below) I will implement a strategy that looks only at schools that changed their schedule during the period of analysis. This allows me to take advantage of the exogenous variation in schools that switched schedules in my estimation strategy. In this case, I argue that given that the Ministry of Education mandates that in the process of allocating spots each year a school should give priority to students that are already enrolled in the schools, students in 5^{th} grade and 9^{th} grade are not likely to have enrolled into a specific school in this period because it was "switching" into a particular schedule. In impact evaluation terminology, this means that 5^{th} graders and 9^{th} graders did not seek to be treated, or alter their condition (enrollment in one school) in order to get the treatment (enrollment in a complete school schedule in 5^{th} grade or 9^{th} grade).

The following sub-sections will describe different identification strategies that address the endogeneity concerns in detail. I argue that by using these strategies, this study mitigates many of the endogeneity problems.

i. Estimation Strategy 1:OLS

A. Cross-sectional OLS

If parents with certain observed characteristics systematically select schools that have a particular schedule, or if the Secretary of Education systematically assigns school spots of a particular schedule based only on observed parental or student socioeconomic characteristics, it may be possible to obtain an unbiased estimate of the impact of attending a complete schedule school with an Ordinary Least Square (OLS) regression at the school level, which controls for the observable characteristics on which there is selection. The variable of interest is COMPLETE, a dummy variable equal to one if the school has a complete (full-day) schedule and zero if does not. It compares schools that have a complete schedule with those that have a morning or afternoon schedule in year t, controlling for observable characteristics and for school performance in the previous period. The following specification is run separately for each grade and subject:

$$\mathrm{SABER}_{\mathrm{st}} = \beta_0 + \beta_1\mathrm{COMPLETE}_{\mathrm{st}} + \beta_2\mathrm{SABER}_{\mathrm{s(t-1)}} + \beta_3\mathrm{X}_{\mathrm{st}} + \mathrm{d}_{\mathrm{m}} + \varepsilon_{\mathrm{s}} \quad (1)$$

Where $SABER_{st}$ is the average test score in SABER for school *s* in year *t* (2005,2009); $COMPLETE_s$ 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 scores in the previous period; X_{st} is a vector of school control variables, like the schools' location (rural/urban), educational level of teachers, and the schools' socioeconomic status; d_m are municipality fixed effects; and ε_s is the error term.

B. Panel OLS

In addition to running the cross-sectional OLS specification, I also run a panel OLS specification. Again in this case, the OLS strategy allows controlling for observable characteristics that could cause selection. However, in this case, I exploit the panel data by including a time dummy variable, to control for any other observed or unobserved policy changes or events that affected all schools equally in each year, and municipality fixed effects, to control for observed or unobserved characteristics that are particular to each municipality. The following is the panel OLS specification:

$$\mathrm{SABER}_{st} = \beta_0 + \beta_1 \mathrm{COMPLETE}_{st} + \beta_2 \mathrm{X}_{st} + \mathrm{d}_t + \mathrm{d}_m + \varepsilon_{st} \quad (2)$$

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 morning or afternoon schedule); d_t are time fixed effects; and all other notation represents the same as in (1).

ii. Estimation Strategy II: School fixed effects model

The problem with the OLS strategies is that it is possible that the observed differences between the two types of school (with a complete, or morning or afternoon scheduler) simply reflect pre-existing differences between them, or their students, and not the impact of attending a particular type of school-schedule. Therefore, I implement a school fixed-effects model to begin to address this problem. School fixed effects allow comparing each school with itself at another point in time when it had a different schedule, controlling for all time-invariant school characteristics that might bias the cross-sectional OLS estimations. For this reason, this specification decreases the bias from observed and unobserved characteristics that do not change over time, although, as I discussed in sub-section iii, there might be time-varying characteristics that might still cause selection issues.

In order to implement this approach, I exploit the fact that some schools switched from having a morning or afternoon shift to having a complete schedule, or vice versa. In this case, the identifying assumption is that schools that switched from having a complete schedule to a single shift, or vice versa, did so for reasons exogenous to the school or to test scores or to unobservable characteristics that are correlated with test scores. Switching into a different type of schedule is a municipality-level decision, and therefore exogenous from the point of view of the school, and the parents or students. The following is the specification for this model:

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

Where d_s are the school fixed effects, and all other notation represents the same as in (1).

iii. Estimation strategy III: Instrumental Variables

OLS model (1) allows controlling only for the observed characteristics that might be correlated to student test scores. OLS model (2) contains time and municipality fixed effects, which allow controlling for characteristics that "hit" all schools in the same way in each time period, and that do not vary in time within a municipality.

The school fixed-effects model (3) allows controlling for any unobservable time-invariant characteristics of the school. However, it cannot capture any time-varying characteristics that could be correlated with both the scheduling switch and student test scores, and therefore might bias the estimations of the impact of having a complete schedule. For example, if in anticipation of a change in school schedule, parents with certain characteristics systematically decide to move their children into a school with a different schedule the year before the change, then changes in student characteristics could drive observed changes in test scores. Alternatively, a change in the school principal correlated with the change in schedule could also be problematic. Examples are if the municipality systematically assigns "better" principals to the schools that are switching to a complete schedule, or if changes in school principals are somehow causing the school to change to a different schedule. If this is the case, then COMPLETE is an endogenous predictor of student achievement, and the school fixed effects model will produce biased estimates of the impact of longer schools days, possibly under- or over-estimating the impact of a change. For this reason, to further address the endogeneity concerns I also use an instrumental variable strategy.

To implement this approach, one needs to find a variable, "the instrument," that is correlated with the potentially endogenous variable (COMPLETE), but that is exogenous to all other factors that might affect student test scores. The instrumental variable is used to predict the probability of having a complete schedule for each school, and then, in a second stage, the outcome is regressed against the predicted values of the variable of interest (COMPLETE). Because of all these characteristics, finding an Instrumental Variable (IV) can be very challenging.

I use the difference between the demand and supply of school spots in each municipality in the previous year to instrument for the probability of having a complete schedule. The identifying assumption in this case is that in municipalities where the supply is larger than the demand for school spots, there is a larger probability that parents will get a spot in the school of their choice. But in municipalities where the demand is greater than the supply for school spots, the Secretary of Education will be more constrained in the school spot allocation process and therefore will assign spots with less consideration of parental preferences. In this case, parents are less likely to systematically select a particular type of school schedule, because students will be assigned to wherever there is a place available in the municipality. There is a possibility that they might get the school of their choice, or their second choice, or another school close to their place of residence, or any other school that has a spot available. One potential problem with this instrumental variable approach is that it is not possible to observe what percentage of parents got their first or second school choices, or what percentage got a school that was not in their two preferences. Therefore, I assume that the larger the gap between the demand and supply for spots in the municipality in year *t*, the less likely it is that parents got their children into their preferred school, and therefore the less likely it is that they systematically selected a particular type of schedule.

To estimate the demand for school spots in each year, I use as a proxy variable the population in the 5th grade or 9th grade age cohort in the municipality in the previous year from the 2005 census. I use the supply of public schools spots in each year by using data from the National Statistics Office (DANE) on the number of enrolled students in 5th grade and 9th grade in the municipality in the previous year (this data comes from consolidating the data on the C-600 school surveys.)

I argue that this might be a good instrument, because it is correlated with the probability of having a complete schedule, but is exogenous to student test scores. First of all, the instrument is likely to be strong because legislation and rules from the Ministry of Education determine that the demand and supply for school spots in the previous year influence decisions on the adjustments necessary to meet the demand. Therefore, the instrument is correlated with the probability of a school having a complete schedule, an issue I will test empirically. As mentioned above, the municipality usually decides on the type of schedule that each school-schedule will have based on the demand and supply for schools spots, so it seems likely that the gap between demand and supply will be correlated with a the probability that a school has a complete schedule or a double-shift: if the demand is higher than the supply in the municipality, then schools in that municipality are more likely to have a double shifts schedules in order to be able to accommodate that large demand. Second, the instrument is arguably valid because the gap between the demand and supply of public school spots is not correlated with test scores or with any other factors that might affect test scores (i.e. with the error term). On one hand, the proxy that I use for demand, the population in the 5^{th} grade or 9^{th} grade age cohort in the municipality, is exogenously determined. And by virtue of their arguably random location in a municipality with a determined age size cohort, schools in that municipality have somehow exogenously "received the treatment" (higher probability of having a particular schedule).

The instrumental variables specification is the following:

First stage: $$COMPLETE_{st} = \beta_0 + \beta_1 (D - S)_{m,t-1} + \beta_2 X_{st} + d_m + d_t + \varepsilon_{st} \quad (4a)$$

Second stage: $$SABER_{st} = \beta_0 + \beta_1 \widehat{COMPLETE}_{st} + \beta_2 X_{st} + d_m + d_t + \varepsilon_{st} \quad (4b)$$

Where *D* and *S* are the demand and supply for educational spots in municipality *m*, and for the second stage, $COMPLETE_{st}$ (with a hat) is the predicted probability of having a complete schedule that comes from the first stage.

I plan to run two variations of the Instrumental Variables model. First, I will run the model at the municipal level: I will use percentage of complete schedule schools in the municipality as the dependent variable in the first stage (4a), and its predicted values as predictors in the second stage (4b). Second, I will try using some variations of the instrument. One possibility proposed by Bonilla-Mejia (2011) is to instrument the variable COMPLETE with the supply of complete schedule schools i.e. the amount of schools with complete schedule in the municipality. The identifying assumption in this case is that the larger the amount of schools with complete schedule in the municipality, the higher the probability of attending a school that has a complete schedule, but the number of schools with a complete schedule in the municipality is not directly related to the school-schedule's student test scores.

IV. Data

To implement the strategies described above, this study exploits an extensive and relatively new panel dataset which contains test score results from SABER 5th and 9th for three periods: 2002-2003, 2005-2006, and 2009. SABER 5th and 9th are nationwide standardized assessments administered every three years to all students in 5th grade and 9th grade in Colombia[4]. 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 competences in students enrolled in basic education, as an indicator of the quality of education[5]. 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.

[4] The 2012 application of SABER was extended to evaluate third grade students, but data for this round is not included in this dataset.
[5] Own translation from the Ministry of Education's website. Retrieved from: http://www.mineducacion.gov.co/1621/w3-article-244735.html, on June 1st, 2013.

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 competences (i.e. abilities) in language (Spanish), math, and science[6], 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.

The SABER panel dataset used in this study comes from the ICFES' databases. Initially, the three rounds of SABER included in this dataset were not comparable. The methodology for implementing the tests changed in the three rounds, and therefore, it was not possible to find information about performance that was comparable through time. However, since 2008, the ICFES, together with the World Bank, has worked in reclassifying the tests administered in 2002-2003 and 2005- 2006, 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 2002-03, 2005-06, 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 during 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 form used annually by the Departamento

[6] The science assessment was introduced in 2012, so it is also excluded from this dataset.

Administrativo Nacional de Estadistica (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 repetitions and approvals in the previous academic year, number of groups/classes by grade, administrative staff, and some information about teachers (for example, 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.

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[7]. 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, morning, or afternoon schedule are very uncommon (mostly night and weekend schedules, and about 5.5% of all schedules-level observations), and their student characteristics might be so different to those of students attending "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 at the student level, but given that the data is not representative at the student level, and that I only have school level characteristics and not

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

student level characteristics, I also 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 or subjects, so I create four different samples within the school panel: 5th grade language, 5th grade math, 9th grade language, and 9th grade math. Each of these samples contains the school's SABER test scores for that grade and subject and its corresponding school characteristics from the C600.

To be part of the panel, a school has to have information from SABER and C600 in at least 2 years. But it is important to note that when matching the SABER and C600 dataset there are some schools that cannot be matched at all.[8] If the observations that are matched and make it into the panel are systematically different from those that are left outside the panel, or to those in the complete SABER or C600 panel before matching, estimations can be biased, and the results may not be consider representative of all schools in the sample. To address this problem, Table 1 contains descriptive statistics from some of the main variables for two different samples: the sample with observations that are left outside the panel, and the sample with the observations that are successfully matched and make it to the panel. For space reasons I only show the sample for 5th grade language, but the other three samples behave in a similar way.

With respect to the SABER dataset, there is only a relatively few number of schools that are left outside the panel. The average test score is very similar between schools inside the panel and the schools in the complete dataset, while the observations left outside the panel have a

[8] There is a significant number of schools in the C600 that cannot be match with any school in the SABER dataset. Although in some cases this is due to problems with matching the school codes in the two datasets, it is also possible that many of the schools in C600 did not have students taking the SABER tests. However, the group of schools that did take the SABER tests, and therefore "exists" in the SABER dataset, can be considered representative of the whole population of schools.

slightly higher average test score in 2002-2003 and 2005-2006, and lower test scores in 2009, than those inside the panel, but in both cases these differences are fairly small. The average socioeconomic status (SES) [9] is lower for the schools inside the panel, mainly because of the much lower percentage of richest schools and higher percentage of poorest schools in that sample. However, the average socioeconomic status is very similar between schools inside the panel and schools in the complete SABER sample.

Regarding the C600 data, there is a substantial amount of schools that are left outside the panel, possibly because of the reasons explained above. The average school enrollment and the average 5th grade enrollment are higher for those schools inside the panel, than for those outside the panel or in the complete C600 panel. With respect to teacher training, the sample of schools inside the panel has a higher percentage of teachers with a professional degree and with pedagogic training, than those schools outside the panel or in the complete C600 panel. In contrast, the teacher student ratio is very similar across all of the samples.

Variables

The dependent variable in this study is the average SABER test score for each grade, subject, and year, for each school. It results from averaging tests scores of all the students that took each of the tests in a school in that year. Following most of the literature that uses the same type of dependent variable, 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

[9] SES is a school socioeconomic status scale included in the SABER dataset, where 1 corresponds to the poorest schools, and 4 to the richest schools.

one. This allows me to interpret the coefficients in the regressions as standard deviations from the mean.

The key variable in this study is an indicator for the type of schedule that a school has in each year: COMPLETE is equal to one if the school has a complete schedule in a year, or zero if it does not[10]. Because other uncommon type of schedules are not included in the panel, 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)[11]. In some 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 (1140 observations, or 0.6% of the sample). However, as a robustness test, I will include these observations and "correct" them by: reclassifying as having a "complete" schedule those schools that reported a morning schedule but NOT an afternoon schedule, and reclassifying as having a morning schedule those schools that reported a complete schedule AND also an afternoon schedule. This method is what the ICFES uses to reclassify the schedule variable in at least one of its SABER datasets.

All of the models and specifications include school characteristics as control variables. These are represented by the vector X in equations (1), (2) and (3). The main school variables included are: a dummy for location in a urban area; a dummy for public schools; school

[10] 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.

[11] However, an overwhelming majority of schools with double shifts has a morning shift for primary and an afternoon shift for secondary. Therefore, when comparing schools that have a complete schedule to schools that do not have a complete schedule in one of the grade-subject samples, in practice I am comparing schools with a complete schedule with schools with a morning or afternoon schedule. For example, schools in the 5th grade language sample that have a complete schedule are compared to schools that have a morning OR afternoon schedule where 5th grade student attend, or to schools with double shifts where 5th grade student attend either a morning or afternoon schedule.

socioeconomic status (SES) (a scale included in the SABER dataset where 1 corresponds to the poorest schools, and 4 to the richest schools); total school enrollment; primary teacher-student ratio; the percent of teachers with a professional education; and the percent of teachers that have specific pedagogic training. Table 2 reports the descriptive statistics for the main variables used in the analysis by type of schedule, for the 5th grade language sample. The other samples are not reported due to space restrictions, but overall, their descriptive statistics seem to follow the same pattern than this sample. Roughly one third of observations for each year corresponds to schools with a complete schedule. For 2005-2006 and 2009, the average test score is higher for schools that have a complete schedule, while schools that have a complete schedule seem to do worse than those with a half schedule in 2002-2003. All of these differences are statistically significant. However, given that schools that have different schedules also differ in many other ways, it is possible that these differences are only reflecting the associations between other characteristics and achievement in SABER. The percent of schools with a half schedule that are located in urban areas is much higher than for schools with a complete schedule. For example, for 2002-2003, 31.6% of schools with a half schedule are located in urban areas, while only 23.5% of schools with a complete schedule are. This reflects the fact that half schedules are more common in urban areas, where schools have to "be split" into different shifts in order to meet the large demand for school spots. Most of the schools in the dataset are public schools, but a larger percentage of schools with a half schedule are public. Schools with complete schedules have a slightly better socioeconomic status (i.e. higher SES), but the difference is only statistically significant for 2009. Total school enrollment is lower in schools with a complete schedule. The difference between teacher student ratios is very small, but statistically significant. Finally, for 2002-2003 and 2005-2006, schools that have a complete schedule have a lower percentage of teachers with a professional education and a lower percent of teachers with pedagogic training. For 2009, schools that have a complete schedule or a half schedule have about the same percent of teachers with professional or pedagogic training, but these differences are not statistically significant.

V. Results

Table 3 presents the results of the cross-sectional OLS model for 5th grade language test scores. This model exploits the variation of schedule and other variables across schools, to estimate how much of the variation in school test scores can be explained by having a complete schedule, controlling for other observable school characteristics. Specification (1) is a basic specification with very few controls. It shows that there is a statistically significant difference in 5th grade language test scores between schools that have a complete schedule and schools that do not: Schools with a complete schedule have test scores that are about one third of a standard deviation (0.050) higher than schools that do not have a complete schedule, controlling for previous performance, location of the school (i.e. urban or rural), and type of school (i.e. public or private). This specification shows that all of these controls are also significantly associated with higher language test scores: a one standard deviation increase in test scores in the previous period is significantly associated with a 0.112 standard deviation increase in test scores in the current year, and language test scores differ by 0.042 of a standard deviation on average between schools located in rural and urban areas, and 0.706 of a standard deviation on average between private and public schools.

The coefficients in specification (1) are likely biased because each of the school characteristics included in the regression can be related to other observable and unobservable characteristics that are associated with school achievement. Therefore, specifications (2) to (6) add one more control at a time. COMPLETE is still statistically significant when adding these controls and remains at about the same levels across these specifications. Additionally, by adding more controls, the location of the school goes from being positively associated with test scores in specification (1), to being negatively associated with test scores in the rest of the specifications. This possibly happens because the difference in tests scores between urban and rural schools might be related to other observable and unobservable differences between rural and urban schools, for example, the school socioeconomic status, or teacher quality. Therefore, once I control for additional observable characteristics, being located in an urban school is associated with lower test scores.

For space reasons, I do not present here the cross-sectional OLS results for 5th grade math test scores, or 9th grade language or math test scores, but they are all consistent with the findings for 5th grade math.

Table 4 presents the findings for 5th grade test scores by subject, using the identification strategies described in section III. The OLS specification (columns 1 and 4) is the cross-sectional OLS model described in equation (1), which also corresponds to the fully-loaded model in Table 3 (column 6) . The panel specification (columns 2 and 5) exploits the availability of multiple observations within school (i.e. data for various years), as described in equation (2). The “school fixed effects” (columns 3 and 6) is the baseline school fixed effect model described in equation (3). The first row presents the estimates for COMPLETE, the impact of having a complete schedule. COMPLETE is consistently statistically significant at the 5 percent level across the three different specifications.

Assessing the language results, the estimate of COMPLETE in the panel specification is higher than for the other two specifications. In this case, having a complete schedule increases language test scores by about one tenth of a standard deviation, compared to schools without a complete schedule. However, as discussed in section III, this estimate could be bias because there might be unobservable differences between schools that have complete schedules and schools that do not. Therefore, the school fixed effects specification is the most reliable of all, because it does not compare differences in student achievement between schools that have different schedules, but compares a school with itself in another point in time. And by including the schools fixed effects, I am able to control for both observable and unobservable time-invariant schools characteristics that could bias the estimates for COMPLETE.[12] In this specification (column 3), the estimate of COMPLETE (0.068) is statistically significant at the 1% level, a little higher than in the OLS specification, but lower than in the panel specification, suggesting that, at least for 5th grade language test scores, having a complete schedule has a larger job in explaining test score variation *between* schools than *within* schools. The estimate for COMPLETE in this specification can be interpreted as the impact, within a school, of having a complete schedule on language test scores.

Columns (4), (5) and (6) in Table 4 show the set of findings for 5th grade math. As with language, the three specifications show that having a complete schedule raises test scores, and that the difference in results between different school schedules is statistically significant at the

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

1% level. For the OLS and panel specifications (columns 4 and 5), the coefficients of COMPLETE are similar to those for language test scores. However, the school fixed effects specification (column 6) shows that math test scores are one tenth of a standard deviation higher on average when schools have a complete schedule, a larger impact of having a complete schedule than for language test scores.

Table 5 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 morning or afternoon schedule, and this is consistent across all specifications, and both for language and math test scores. All the coefficients for COMPLETE in these specifications are larger than those for 5th grade, suggesting that having a complete schedule has a larger positive impact on 9th graders than on 5th graders. For example, the coefficients for COMPLETE in the OLS specifications (columns 1 and 4) indicate that having a complete schedule is associated with an increase of almost one fourth of a standard deviation in language and math test scores, about four times the size of the coefficients for the 5th grade regressions reported in columns 1 and 4 of Table 4.

The findings from the panel specifications also show a substantially higher impact of having a complete schedule on both language and math test scores for 9th graders. While having a complete schedule raises test scores by about one tenth of a standard deviation for 5th graders, it raises test scores by about one third of a standard deviation for 9th grade language test scores, and by 0.289 of a standard deviation for 9th grade math test scores.

The school fixed effects specifications shows that the within school impact of having a complete schedule is twice as large for 9th grade language test scores than for 5th grade language test scores, although the impact is relatively more similar for 5th grade and 9th grade math test scores (0.102 v. 0.145). On the other hand, similar to the results for 5th grade test scores, the estimates for COMPLETE for 9th graders are lower in the school fixed effects specifications than in the panel specifications. For example, the estimated impact of having a complete schedule for language test scores in the panel specification (column 2) is about three times as big as the impact estimated in the school fixed effect specification (column 3). The same number is about twice as big for the math test scores (columns 5 and 6). Part of the reason for this differences is that the school fixed effect model might be a better estimation of the "causal" impact of having a complete schedule, because most confounding factors that could potentially bias the estimations for COMPLETE originate from differences between schools. This result also strengthens the idea that the differences in schedules explain relatively more of the between school variation than the within school variation. Finally, the school fixed effects results for 9th grade (columns 3 and 6) show that the impact of having a complete schedule is fairly similar for math and language test scores.

To sum, the most reliable estimates of the impact of having a complete schedule come from the school fixed effects specifications. These estimates range from 0.068 of a standard deviation for 5th grade language test scores, to 0.145 of a standard deviation for 9th grade math test scores.

Results from the Instrumental Variables model are still preliminary and are not reported in this version of the paper.

VI. Conclusions

This study provides some of the first evidence that longer school days has an impact on 5th grade and 9th grade academic achievement in Colombia. By exploiting changes in the type of schedule, I am able to use a school fixed effect model to control for observed and unobserved time-invariant characteristics, mitigating in this way many of the endogeneity concerns. I find that there is a positive impact of having a complete schedule (approximately 2-3 additional hours) on schools achievement in 5th grade and 9th grade SABER test scores. Results from the school fixed effects model show that schools with a complete schedule have about one tenth of a standard deviation higher test scores than those with half schedules. To put the magnitude of this effect in perspective, this effect is similar to moving one standard deviation up the teacher quality distribution (Rivkin, Hanushek, and Kain, 2005) or reducing class size by 4 students (Krueger, 1999). The impact is larger for math test scores than for language test scores in both grades, and it is stronger for 9th grade students than for 5th grade students. These results suggest that having longer schools days might be a policy with a larger impact for 9th graders, perhaps because by keeping them in school for longer, they stay “out of trouble” and spend more time studying, which ultimately translates into better academic achievement.

The findings of this study complement those of previous studies that have found a positive impact of attending a complete schedule on early dropout and grade repetition (Garcia, Fernandez and Weiss, 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, and therefore increasing the length of the school day is a strategy worthwhile to consider in order to improve the quality of education in developing countries.

VII. References

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., Bertrand, M., Linden, L., and Perez-Calle, F. (2011). Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia. *American Economic Journal: Applied Economics, 3*(2*),*167-195.

Berliner, D. (1990). What's all the fuss about instructional time? In Ben-Peretz, M. and 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*),* 629-40.

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

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

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

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

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

Garcia, S., Fernandez, C., and 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. and Wossermann, L. (2007). Education quality and economic growth. World Bank. Washington, DC.

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., and 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. (2000). The effect of class size on student achievement. New evidence from population variation. *Quarterly Journal of Economics, 115*(4*),* 1239-85.

Hoxby, C., Murarka, S. (2009). Charter school in New York City: Who enrolls and how they affect their students' achievement. NBER Working Paper 14852.

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

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. NBER working paper series No. 16227.

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

Llach, J. ,Adrogue, C., and 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), pp. 629-40.

Marcotte, D. and Hansen, B. (2010). Time for School. Education Next. v. 10, n. 1, pp. 52-59

Margo (1994). Explaining Black-White Wage Convergence, 1940-1950. Industrial and Labor Relations Review, Vol. 48, No. 3 (Apr., 1995), pp. 470-481

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

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

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.

Rivkin,S., Hanushek, E. and 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 data inside and outside the school panel. Sample: 5th grade language.

	2002-2003		2005-2006		2009	
	Outside panel	**Inside panel**	**Outside panel**	**Inside panel**	**Outside panel**	**Inside panel**
No. of schools	535	18563	825	19359	1561	31781
Average test score	298.81	297.28	289.65	288.71	280.11	289.55
Mean standard deviation of test scores	46.47	53.97	43.64	44.22	56.75	54.69
Mean SES	2.23	1.69	1.86	1.74	1.63	1.65
% of poorest schools (SES= 1)	34.93	63.2	54.38	61.91	64.57	64.48
% of richest schools (SES=4)	21.5	12.02	14.43	13.03	9.65	10.78
Average school enrollment	128.63	224.67	109.85	270.59	99.98	205.56
Average 5th grade enrollment	17.31	23.27	14.85	28.96	12.78	23.06
Primary teacher/student ratio	0.05	0.04	0.06	0.05	0.08	0.06
Primary teachers w/professional degree	41.51%	57.19%	54.62%	65.39%	47.92%	64.25%
Primary teachers w/pedagogic training	43.96%	59.13%	55.12%	65.52%	49.67%	64.92%

Table 2. Descriptive statistics of school panel by type of schedule. Sample: 5th grade language.

	2002-2003			2005-2006			2009		
	Half (SD)	**Complete (SD)**	**Diff. (t-stat)**	**Half (SD)**	**Complete (SD)**	**Diff. (t-stat)**	**Half (SD)**	**Complete (SD)**	**Diff. (t-stat)**
Mean test score	0.03	-0.05	**0.08**	-0.08	0.10	**-0.18**	-0.06	0.13	**-0.19**
	(0.98)	(1.02)	**(4.65)**	(0.95)	(1.05)	**(-11.23)**	(0.97)	(0.99)	**(-15.50)**
% of urban school	31.6%	23.5%	**8.1%**	41.0%	27.0%	**0.14**	31.6%	24.4%	**0.07**
	(0.46)	(0.42)	**(10.89)**	(0.49)	(0.44)	**(19.36)**	(0.47)	(0.43)	**(12.84)**
% of public schools	88.3%	80.4%	**7.9%**	89.5%	78.1%	**0.11**	91.9%	79.2%	**0.13**
	(0.32)	(0.40)	**(12.34)**	(0.31)	(0.41)	**(19.32)**	(0.27)	(0.41)	**(26.94)**
Mean school SES	1.67	1.70	-0.03	1.73	1.74	-0.01	1.62	1.73	**-0.12**
	(1.00)	(1.15)	(-1.44)	(1.02)	(1.19)	(-0.58)	(0.94)	(1.17)	**(-8.24)**
Mean school Enrollment	251.79	158.07	**93.72**	334.24	162.81	**171.43**	243.21	147.95	**95.26**
	(412.18)	(275.64)	**(17.1)**	(486.51)	(277.57)	**(29.72)**	(416.93)	(277.37)	**(22.96)**
Mean teacher student ratio	0.04	0.04	**0.00**	0.05	0.05	0.00	0.06	0.07	**-0.01**
	(0.03)	(0.02)	**(-2.69)**	(0.18)	(0.06)	(-1.38)	(0.19)	(0.16)	**(-3.25)**
% of teachers w/professional education	59.0%	55.3%	**3.7%**	67.9%	62.8%	**0.05**	64.7%	64.6%	0.00
	(0.40)	(0.42)	**(5.18)**	(0.37)	(0.40)	**(8.34)**	(0.40)	(0.41)	(0.19)
% of teachers w/pedagogic training	60.2%	58.4%	**1.9%**	67.7%	63.5%	**0.04**	65.3%	65.1%	0.00
	(0.39)	(0.42)	**(2.62)**	(0.37)	(0.40)	**(6.75)**	(0.40)	(0.41)	(0.38)
Observations (No. of schools)	11850	4842	-	10994	6561	-	20256	8915	-

Note: Statistiscally significant differences are bolded.

Table 3. OLS regressions. Dependent variable: 5th grade language test scores

	(1)	(2)	(3)	(4)	(5)	(6)
Complete	0.050*	0.053**	0.056**	0.055**	0.055**	0.056**
	[0.028]	[0.026]	[0.027]	[0.027]	[0.026]	[0.026]
Test score in previous period	0.112***	0.105***	0.105***	0.104***	0.103***	0.103***
	[0.010]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Dummy for urban schools	0.042**	-0.126***	-0.147***	-0.147***	-0.155***	-0.156***
	[0.021]	[0.026]	[0.030]	[0.030]	[0.030]	[0.029]
Dummy for public schools	-0.706***	-0.497***	-0.532***	-0.535***	-0.572***	-0.572***
	[0.032]	[0.040]	[0.033]	[0.033]	[0.032]	[0.032]
School socioeconomic status		0.184***	0.178***	0.178***	0.174***	0.174***
		[0.023]	[0.022]	[0.021]	[0.021]	[0.021]
School enrollment			0.006**	0.005**	0.005**	0.005**
			[0.002]	[0.002]	[0.002]	[0.002]
Primary teacher-student ratio				-0.041**	-0.038*	-0.037*
				[0.020]	[0.020]	[0.020]
% teachers with professional education in primary					0.107***	0.082***
					[0.016]	[0.030]
% teachers with pedagogical training in primary						0.030
						[0.029]
Observations	32,711	32,711	32,711	32,711	32,711	32,711
R-squared	0.078	0.089	0.089	0.089	0.091	0.091
Number of municipalities	1,037	1,037	1,037	1,037	1,037	1,037

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

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 is an index from the SABER data that goes from 1 (poorest schools) to 4 (richest schools).

Table 4. Regressions results. Dependent variable: 5th grade test scores.

	Language			Mathematics		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Panel	School fixed effects	OLS	Panel	School fixed effects
Complete	0.056**	0.104***	0.068***	0.060**	0.102***	0.102***
	[0.026]	[0.020]	[0.020]	[0.026]	[0.021]	[0.020]
Test scores in previous period	0.103***	--	--	0.127***	--	--
	[0.008]			[0.009]		
Dummy for urban schools	-0.156***	-0.162***	--	-0.163***	-0.222***	--
	[0.029]	[0.026]		[0.029]	[0.026]	
Dummy for public schools	-0.572***	-0.460***	--	-0.531***	-0.428***	--
	[0.032]	[0.028]		[0.030]	[0.030]	
School socioeconomic status	0.174***	0.162***	--	0.137***	0.114***	--
	[0.021]	[0.017]		[0.019]	[0.014]	
School enrollment	0.005**	0.001	-0.005	-0.001	-0.006	-0.006*
	[0.002]	[0.002]	[0.003]	[0.003]	[0.004]	[0.004]
Primary teacher-student ratio	-0.037*	0.000	0.061**	-0.045**	-0.007	0.012
	[0.020]	[0.019]	[0.025]	[0.019]	[0.014]	[0.022]
% teachers with professional education in primary	0.082***	0.074***	-0.035	0.059*	0.059**	-0.016
	[0.030]	[0.025]	[0.029]	[0.031]	[0.027]	[0.029]
% teachers with pedagogical training in primary	0.030	0.025	0.035	0.048*	0.031	0.031
	[0.029]	[0.023]	[0.028]	[0.029]	[0.024]	[0.028]
Observations	32,711	63,418	63,418	32,836	63,606	63,606
R-squared	0.091	0.046	0.001	0.079	0.035	0.002
Year fixed effects	NO	YES	YES	NO	YES	YES
Municipal fixed effects	1,037	1,093	NO	1,036	1,094	NO
School fixed effects	NO	NO	31,089	NO	NO	31,155

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 is an index from the SABER data that goes from 1 (poorest schools) to 4 (richest schools). In the panel specifications, standard errors are robust. In the school fixed effects specifications, standard errors are clustered at the school level.

Table 5. Regressions results. Dependent variable: 9th grade test scores.

	Language			Mathematics		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Panel	School fixed effects	OLS	Panel	School fixed effects
Complete	0.245***	0.357***	0.122***	0.229***	0.289***	0.145***
	[0.032]	[0.037]	[0.031]	[0.035]	[0.035]	[0.035]
Test scores in previous period	0.222***	--	--	0.175***		--
	[0.030]			[0.034]		
Dummy for urban schools	-0.123***	-0.164***	--	-0.125***	-0.164***	--
	[0.026]	[0.034]		[0.037]	[0.032]	
Dummy for public schools	-0.433***	-0.441***	--	-0.460***	-0.441***	--
	[0.040]	[0.042]		[0.038]	[0.045]	
School socioeconomic status	0.277***	0.314***	--	0.208***	0.193***	--
	[0.018]	[0.023]		[0.022]	[0.020]	
School enrollment	0.015***	0.014***	-0.022***	0.013***	0.011***	-0.013***
	[0.002]	[0.002]	[0.004]	[0.003]	[0.003]	[0.004]
Primary teacher-student ratio	-0.189***	-0.098***	-0.048	-0.122**	-0.048**	-0.132*
	[0.053]	[0.027]	[0.057]	[0.062]	[0.023]	[0.070]
% teachers with professional education in primary	0.088	0.174***	-0.042	0.083	0.131**	-0.090
	[0.055]	[0.065]	[0.056]	[0.070]	[0.065]	[0.060]
% teachers with pedagogical training in primary	0.006	-0.013	-0.052	-0.047	-0.039	-0.023
	[0.049]	[0.039]	[0.053]	[0.056]	[0.043]	[0.057]
Observations	9,289	18,424	18,424	9,288	18,435	18,435
R-squared	0.326	0.200	0.008	0.195	0.110	0.005
Year fixed effects	NO	YES	YES	NO	YES	YES
Municipal fixed effects	1,014	1,083	NO	1,014	1,083	NO
School fixed effects	NO	NO	9,258	NO	NO	9,270

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 is an index from the SABER data that goes from 1 (poorest schools) to 4 (richest schools). In the panel specifications, standard errors are robust. In the school fixed effects specifications, standard errors are clustered at the school level.