The Impact of School Renovation and Modernization on Students' Academic Outcomes

-

The Portuguese experience

João Francisco Belo Campos de Lima

Master Thesis

Erasmus University Rotterdam

Erasmus School of Economics

Supervised by

Prof. dr. H. D. Webbink

University:	Erasmus University Rotterdam
Faculty:	Erasmus School of Economics
Title:	The Impact of School Renovation and Modernization on Students' Academic Outcomes – The Portuguese experience
Author:	João Francisco Belo Campos de Lima
Student Number:	405858
Specialization:	Policy Economics
Date:	November 2015

Abstract

This paper looks into a school renovation and modernization program undergone in Portugal to assess the impact of modernized school infrastructure on students' academic outcomes. It uses a difference-in-differences approach and a fixed-effects approach and estimates a negative impact of the program on student grades both in the period in which schools are experiencing renovation works as well as in the years shortly after these works are finished and the modernized schools reopen to their normal academic functioning.

Key words: schools, renovation, modernization, Portugal, students

Table of Contents

List of Tables

List of Figures

1. Introduction

In light of the present knowledge-based economy and the remarkable expansion of the service sector in developed societies, accumulation of human capital has become a central policy goal for governments and policymakers as a means to enhance labour productivity and economic growth. The OECD (2010), in its report *The High Cost of Low Educational Performance,* has shown that relatively small improvements to labour force skills can largely impact the future wellbeing of a nation, through substantial gross domestic product gains. These findings are only a small part of the large body of evidence contributing to the overwhelming consensus on the positive impact of education on societies, which has led governments to devote huge amounts of resources towards improving the educational standards of their citizens.

While school facilities are among the most popular spending categories for education policymakers, the degree to which improving school infrastructure and learning spaces is effectively meeting the goals of these knowledge-seeking societies is still a matter under research. The present study contributes to this discussion by trying to answer the question of whether the quality of learning environments – which includes the structural quality of school buildings, the design of learning spaces and their technological equipment - are important factors for educational achievement and, if this is the case, estimating the magnitude of this impact.

It is difficult to get credible estimates of the effects of spending on school facilities, as exogenous variation in the quality of school facilities is rare. Investment in these facilities is typically endogenous as schools are free to choose how much to invest in their infrastructure, and they often do so according to the needs of their students. This implies that a comparison of schools that have improved their facilities with schools that did not might be biased by unobserved differences between these two groups of schools. In this study I am able to overcome this challenge by exploiting the variation in spending on school facilities coming from changes in a specific program in Portugal.

In 2007, the Portuguese government created a school modernization program, aimed at renovating most of the secondary school network in the country. While this program failed to reach its goal of covering most of the schools in the country, it was still able to intervene in a significant number of buildings, whilst also leaving a good number schools without intervention. Using a difference-in-differences and fixed effects approach, I am able to isolate the contribution of the program to the evolution of student grades in exams of math, Portuguese, biology, geology, physics, chemistry and others, by comparing outcomes between treated and control schools. Moreover, the existence of a group of schools that was scheduled to participate in the program but, due to a slowdown of the program following the economic and financial turmoil of the last years, ended up not being intervened, allowed for the creation of a more comparable control group, and thus a more accurate impact estimation.

The results of this research point to a negative impact of the program on student grades during the phase of the program in which renovation works are occurring in the schools. This reduction in student performance is especially significant for the science and engineering (S&E) related courses: math, biology, geology, physics and chemistry. In the years shortly after the renovation works have ended, student grades in modernized schools continue to be negatively affected. While further research at a later date is needed to assess the real causes of this negative impact, this research will equate some possible explanations.

2. Literature Review

There is a large body of research investigating the link between the quality of learning environments and student achievement. Most of this research is very diverse in regards to the aspects of the "learning environment" which are considered.

A first body of literature is related to additional expenditure in schools, through funding schemes in which some schools obtain more resources than others to invest in their learning spaces as they see fit. These studies find mixed results of the impact of extra resources for schools on student outcomes (see Guryan, 2000; Papke, 2005; Card and Payne, 2002; Chay et al., 2002; Leuven et al., 2007). Other strands of literature focus on the impact of additional investment in information and communication technology (ICT), by assigning computers and other equipment to a few selected schools and comparing their outcomes with control schools, usually finding no effect of increase ICT spending (see Angrist and Lavy, 2002; Goolsbee and Guryan, 2006; Leuven et al., 2007; Machin et al., 2007). Finally, another body of research focuses on the impacts that investing in school facilities has for the outer community, by looking at its relationship with housing prices in the school's surrounding areas (Cellini et al., 2010).

While all these branches of research might be of indirect interest to the present paper, none of them directly answers this paper's research question: how does, specifically, improving the structural, environmental and design quality of school buildings contribute to student outcomes? The "effective school" literature focuses on this specific research question, and while most of it is not as accredited as the research cited above, it has produced some interesting findings.

Most of the "effective schools" literature carried out in the United States is of quantitative nature. In Europe, qualitative studies derived from the social science methodology are the norm (Fisher 2001). Most of the quantitative research points to a five to eight percent increase in academic achievement for students in better condition facilities. For example, Plumley (1978) has found poorer results in assessments of basic skills for students of non-modernized school buildings in the state of Georgia, United States. Bowers and Burkett (1987) found a similar impact on student outcomes as well as a better record for students in modernized schools in the areas of health, attendance, and discipline. Thomas (1962), Chan (1979), McGuffey and Brown (1978) and Frazier (1993) have all concluded there is a positive causal relationship between a better built environment and student outcomes. Gerritsen (2014) has found a positive impact of improving infrastructure for science and engineering related courses in student enrolment in these areas. While new students of lower ability were attracted to these courses, student outcomes did not decrease as a result, suggesting that the improved infrastructure might have had a positive contribution to student achievement.

Some studies have focused on more specific factors that contribute to the quality of school buildings. These building conditions can be divided into two categories: the *structural* conditions, which may include, among others, the floor and ceiling materials, the heating, lighting and noise conditions; and the *cosmetic* conditions, which include for example the painting of walls, the furniture, or the landscaping. Research work by Tanner and Jago (1999), Cohen et al (1986), Earthman (1998), Sinofsky and Knirck (1981), and Buckley et al (2004) have linked higher student performance with improved conditions of air movement and ventilation, thermal comfort, classroom lighting, natural daylight and acoustics. By contributing to the aesthetic and psychological character of learning spaces, these factors were found to increase student efficiency, work output, performance and the attention span.

A much more recent area of research explores the impact of the design and flexibility of learning spaces in supporting individual learning, formal and informal learning as well as collaborative and project work (see Shor, 1996 and DEYTA, 1993). Another also emerging body of research focuses on the theory that opening the school to the outer community, by providing of wider community services and activities in schools, has a positive impact on student achievement, attendance and behaviour. As the school becomes regarded as a site of resources and support for the community, parents engagement with the school increases, leading to higher parental support for students and more timely responses from the school to broader family and student needs (Wilkin et al, 2002).

3. The Portuguese Case

Construction of public secondary schools in Portugal began in the late 19th century, with 23% of the total number of schools still open today being built from this date until the 1960s. Of the remainder 77%, some 46% of the schools were built during the 1980s (Parque Escolar, n.d.a). As a result, in the early 2000s, the 511 schools that constituted the public secondary school system in Portugal comprised a heterogeneous network, both in terms of school typological/morphological conditions and their architectural and building quality.

Considerable efforts were made over the last four decades in terms of creating school infrastructures, but these focused mainly on expanding the school network, and thus not providing for "the parallel development of a constant and consistent practice in terms of conservation and maintenance of the existing buildings or their functional adaptation to reflect new developments in curricula and new educational and training practices" (Parque Escolar, n.d.b). This had led to diverse signs of physical, environmental and functional degradation in schools, despite their high level of physical robustness, and to a mismatch with educational needs of the modern era.

With the purpose of reversing this process of deterioration of the learning conditions, the Portuguese Ministry of Education created the *Secondary School Modernisation Program* (SSMP). This program aimed at achieving an effective renovation of most public secondary schools, adapting them to the new educational and environmental paradigms and guaranteeing the existence of a management model that can effectively meet, with cost control, normal demands in terms of conservation and upkeep (Parque Escolar, n.d.b). Specifically, the program included the correction of existing construction/building problems, changes in hygrothermics, acoustics, air quality, safety and accessibility, adapting teaching and non-teaching spaces, modernising the respective equipment and providing schools with the spaces needed for their use by the wider local community during after-school periods. Despite this diverse set of interventions, the primary focus was put on renewing the physical infrastructure of schools, as indicated by the fact that 86% of all the investment done with the program was channelled to construction works on school buildings (Parque Escolar, n.d.c).

For all the planning, management, development and execution of the school modernisation programme, *Parque Escolar,* a corporation governed by public law and funded by the government, was created in 2006. At the beginning of its activity in March 2007, Parque Escolar's objectives included the intervention in 332 schools by 2015 (Parque Escolar, n.d.d). The program was to be implemented in five phases: a trial phase (phase 0) with only four schools, which started in 2007, followed by interventions in a much larger number of schools to start in 2009 (phase 1), an extension of the program still in 2009 with more schools being intervened (phase 2), another extension to start in 2010 (phase 3) and a final extension to start at a further date (phase 4). On average, each school was to receive 15 million euros worth of investment (Parque Escolar, 2011).

Given the economic and financial turmoil that affected the economic environment internationally, and its consequences to the Portuguese public finances, the goals of the SSMP, then already underway, were re-evaluated during the second semester of 2011 to include a smaller number of schools and a reduction of the total investment. As a result, some of the schools scheduled to be intervened in phase 3 that were awaiting the beginning of construction saw their modernization processes suspended, and the initial procedures for the launching of phase 4 were not initiated (Parque Escolar, n.d.e). The last publicly available figures point to a total investment realized up until 2011 rounding the 2.405 million euros, as stated in the external audit of the Inspectorate General of Finances (Inspecção Geral das Finanças, 2011).

By request of the Portuguese Ministry of Education, in 2009 the program was subject to qualitative evaluation by the OECD's Centre for Effective Learning Environments, which has described the program as having created "significant achievements" through the "radically improved" quality of the buildings (OECD, 2009).

4. Data

4.1. The outcome variable: Academic Achievement

Information on the evolution of the academic achievement of all Portuguese students, the variable of interest in this research, was gathered from the Portuguese Ministry of Education database, in the form of exam grades (DGE, n.d.a; DGE, n.d.b).

Every year, students of all secondary schools in Portugal (public and private) are required undergo nationwide exams for a set of courses (that vary according to each student's academic *track/*area of specialization), in order to be eligible to get their diploma and apply for higher education institutions. These exams are the same for all schools and are conducted simultaneously (at a certain date and time at the end of the academic year - usually in June), allowing for an unbiased comparison of results. Student-level data on the grades in each of these exams, which are given on a scale of 0 to 200, for all schools in the country (intervened by the SSMP and otherwise) from 1998 to 2015 were compiled, totalling more than five million grades. Specifically, for each school in each year, grades of the following exams were computed: math, portuguese language, biology with geology (simply referred to as "biology" from this point onwards) and physics with chemistry (simply referred to as "physics" from this point onwards). Another set of grades that was gathered

Table 1: Sample of exams - Number of student-level exam grades collected by course and year

Year	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	Total
Math	66,7	69,0	43,6	58,0	81,7	76,2	78,4	83,0	67,5	65,8	54,9	59,5	58,2	66,0	76,6	76,2	74,9	1 156,2
Portuguese	68,1	68,2	66,7	62,7	68,2	66,6	71,2	72,0	50,7	76,2	87,3	84,3	91,5	100,9	96,7	97,1	87,9	1 316,4
Biology	41,6	45,1	47,1	45,2	55,7	55,5	53,5	48,5	32,6	54,3	64,4	63,6	64,3	63,1	73,1	76,5	71,6	955,5
Physics	49,4	50,6	51,0	47,5	56,3	57,1	56,3	56,7	36,5	54,0	60,8	67,4	68,3	66,5	78,1	80,4	76,4	1 013,5
Rest	34,7	33,3	30,9	29,4	33,7	34,6	32,6	32,0	37,7	60,6	67,0	75,5	74,6	76,2	78,8	79,2	77,2	887,9
Total	260,5	266,1	239,3	242,8	295,7	290,0	292,1	292,2	225,0	310,8	334,3	350,4	357,0	372,7	403,2	409,5	388,0	5 329,5

Values in thousands of exams

Table 2: Sample of observations - Number of school-level average grades by course

Year	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	Total
Math	572	542	543	553	555	557	555	563	560	562	586	569	574	585	600	593	618	9 687
Portuguese	540	545	529	556	559	558	560	567	560	564	589	572	579	588	602	606	615	9 689
Biology	523	529	537	547	558	553	553	558	537	530	584	573	579	587	602	609	612	9 571
Physics	523	530	540	548	551	550	551	558	558	561	584	574	579	588	602	609	611	9 617
Rest	537	535	545	547	548	544	544	548	529	567	570	574	579	588	599	593	594	9 541
Total	2 695	2 681	2 694	2 751	2 771	2 762	2 763	2 794	2 744	2 784	2 913	2 862	2 890	2 936	3 005	3 010	3 050	48 105

included grades in the exams of geography, economics, drawing, descriptive geometry and history. This last set of grades will be grouped together and be referred to as *Rest,* as it refers to the rest of the exams that are not going to be evaluated on a stand-alone basis. The number of exam grades collected by course and year is summarized in table 1.

In order to create the final dataset, an average grade for each school, in each exam, in each year, was computed. As a result, each individual observation in the dataset will look like, for example, *schoolN99math,* if referring to the average grade of school N, at year 1999 for the math exam. Table 2 summarizes the observations contained in the final dataset. A total of 48.105 observations were obtained. These average grades are used as a proxy for the educational achievement of each school – the dependent variable.

4.2. The treatment variable: intervention under the SSMP

For each school, having received *treatment* means being selected for and undergoing physical renovation and modernization of their buildings under the SSMP program. To determine which schools did so, a complete list of all secondary schools (intervened by SSMP or otherwise) was compiled, totalling 644 schools, from which 511 are public schools and 133 are private schools. Data on which schools were selected and effectively intervened under the SSMP were compiled from Parque Escolar's websit (Parque Escolar, n.d.f).

Of the 644 schools that constitute the Portuguese secondary school network, *Parque Escolar*'s objectives at the beginning of its activity in March of 2007 included the intervention in 332 of the public schools by 2015. However, only 294 ended up being selected for the programme. The program was to be executed in five sequential phases. Four secondary schools were selected for the trial phase (phase 0), 25 were selected for phase 1, 73 were selected for phase 2, 104 for phase 3 and 88 for phase 4. Due to several delays in the building process and the slowdown of the plan in 2011, in virtue of the economic and financial situation of the country, by August 2015 only 142 schools had already been intervened while other 25 were under intervention. This was a result of the suspension of 35 out of the 104 schools that constituted phase 3, and of all 88 schools of phase 4. The remaining four schools that were selected for the SSMP are schools for which data on their status was not publicly available, hence they will be eliminated from the sample. A summary of these data is presented below, in table 3.

Table 3: Sample of schools by status of intervention under the SSMP

Selected Schools					
Program Phase	Schools *Intervened*			Schools *on-hold**	Schools Selected
	Finished*	U/C* **	No data		
0	4	0	0	0	4
1	25	0	0	0	25
2	68	1	4	0	73
3	45	24	0	35	104
4	0	0	0	88	88
TOTAL	142	25	4	123	294
Not-Selected Schools					
TOTAL					350
				TOTAL	644

* (As of August 2015); ** Under Construction

As a result of the developments in the program, it is possible to distinguish three different groups of schools. The first group, referred to as *intervened schools* from this point onwards, is comprised of the schools that have already undergone the renewal process under the SSMP, regardless of whether they have already reopened their upgraded facilities to the normal academic functioning or they are still under construction. This is the *treatment* group. The second group, which will be referred to as *on-hold schools* from now on, is comprised of all the schools that were selected to be intervened by the SSMP but, for whichever reason, ended up not being intervened as of now. The third and final group, referred to as *not-selected schools* from now on, is comprised of all the schools that were neither intervened by the SSMP nor selected to be intervened. This last group includes some public schools as well as all private secondary schools in the country, which were not a target of this public program. These last two groups, together, form the control group in this research.

4.3. Time Periods

In order to analyse the evolution of academic outcomes in the transition from the period before the SSMP's intervention to the period after the SSMP'S intervention, these time periods have to be defined. However, despite the main focus of this research work being the before and after differences in academic achievement created by the SSMP, the period during which the schools are being subject to construction works may also be interesting to look at. Several schools that have gone through the SSMP have reported a significant worsening of the learning environment during the years in which construction was occurring. In fact, besides the frequent complaints about the overly high levels of noise and functional nuisance created by the construction works ("Ainda há 90 escolas", 2010), in most cases all students were displaced from their usual classrooms and relocated to prefabricated steel containers (modular classrooms), sometimes for several years, with lower acoustic, technological and environmental standards than their old schools (Inácio, 2014). In some cases, the temporary dismantling of the existing natural sciences' laboratories for future upgrading left teachers temporarily unable to conduct all the experimental work required for preparing students for the exams of S&E courses (Prof. A. Costa, personal communication, June 14, 2015). Because all these drawbacks may have had a negative impact on exam grades, especially in the Biology and Physics exams, a *before-construction* versus *during-construction* comparison may also shed some light into the impact of programs like the SSMP in the period of construction.

Therefore, for each school, three time periods were defined, using data directly supplied by Parque Escolar on the starting and ending dates of construction of each of the intervened schools: the period *before*, which refers to the years prior to the SSMP's intervention; the period *during*, which refers to the years during which construction and renovation works carried out by the SSMP were taking place; an the period *after*, which refers to the years after the construction and renovation works carried out by the SSMP are finished and the modernized school re-opens to its normal academic functioning.

The difficulty in defining the *during* and *after* periods is that there isn't a fixed year/time period in which all schools started and completed their modernization processes – each school started and ended their modernizations process at different dates. In absence of a fixed time frame for all schools, a dummy variable *after* was created, and it equals one for all observations that occurred after the end of the construction works for the SSMP. For example, if construction in school N occurred between 2008 and 2010, the variable *after* would take the value of one for the observation schoolN2011math, since the math exam conducted in 2011 in school N was done after the construction works have ended. Following the same reasoning, it will take the value of zero, for example, for observations schoolN2009portuguese or schoolN2007physics. Similarly, a dummy variable *during* was created, which equals one for all observations that took place during the construction period of the SSMP. This variable will enable to capture changes in exam grades caused by the construction process.

In order to explore the connection between the SSMP's intervention in schools and the exam grades, the time periods in which construction took place need to be aligned with the years in which exams took place. This is not a very straightforward matter and requires some methodological choices.

Since exams are conducted in June, we cannot expect to see relevant changes in the grades of an exam of 2010 for a school that went through the SSMP and re-opened after the exam, for example, in October 2010. Despite re-opening in 2010, it makes no sense that, for this school, the variable *after* takes the value of one in 2010. Even if a school re-opened before the exam of June, if it did so only a few months before (lets say it re-opened in April), it is not plausible that the grades in June will already reflect the effects of the improved infrastructure, as students have only benefited from them for the short period of two months, out of the nine months that constituted that academic year. Hence, it becomes pointless to have the variable *after* take the value of one in 2010 for all schools that reopened in 2010, even if they re-opened before the exam of that year. Instead, what is likely to be reflected in the exam of June of 2010 is the impact of the construction works that have been going on in the months up to the re-opening, thus the variable *during* is the one that must take the value of one in 2010.

As such, in the models of this paper the rule will be that the variable *after* will take the value of one in a certain year if the school's reopening date occurred at least six months before the exam of June of that year. For example, *after* will take the value of one in 2010 for all schools that reopened before January of 2010, and *during* will take the value of one for all schools that opened after that date.

The same methodology applied to draw the line between the variables *after* and *during* applies when separating the variable *during* from the period *before*. If a school begins construction in May 2008 this may not impact the exam preparation of students too much for it to be visible in the exam grades one month later. Hence, the variable *during* will take the value of one only when schools started construction more than six months before the exam of June. Thus, for this school, the variable *during* will only take the value of one later in 2009, while in 2008 it's the *before* variable that will take the value of one. Later in the sensitivity analysis, this rule of the six months will be relaxed to allow for different time delays.

Table 4 summarizes the sample of observations when these time periods are taken into account.

Table 4: Sample of observations by status of intervention under the SSMP and time period

		Treated Schools		
		Av. Grades		
	Before	During	After	TOTAL
Intervened	10 334	1 682	2 039	14 055
		Control Schools		
		Av. Grades		
	Before	During	After	TOTAL
Not Selected	24 010	0	0	24 010
On-hold	10 040	0	0	10 040
			TOTAL	48 105

4.4. Socioeconomic Data

In most studies that evaluate the impact of investments in schools on student outcomes, variation in the quality of school facilities is endogenous as schools are free to choose how much to invest in their infrastructure, and they often do so according to their students' needs. This implies that a comparison of schools that have improved their facilities with schools that did not might be biased by unobserved differences between these schools.

The most common source of bias is the socioeconomic levels of the students, defined as the social standing or *class* of individuals or groups, that derives from a combination of education, income and occupation attributes. As demonstrated by an extensive body of research, children from communities and households of a lower socioeconomic status (SES) develop academic skills more slowly compared to children from higher SES groups (Morgan, Farkas, Hillemeier, & Maczuga, 2009). Therefore, the confounding effect of socioeconomic status on the impact of modernized school buildings can be a twofold. On the one hand, schools with students from a lower SES have a higher need for improving learning environments, thus they might invest more in school infrastructure than schools with students of higher socioeconomic backgrounds. On the other hand, schools with students of a lower SES are usually located in regions that are relatively more underdeveloped than regions with students of a higher SES. If this has repercussions on the amount of investment each government is able to spend on their public schools, then one would expect schools of wealthier regions (which also have students of higher SES) to receive more investment (Aikens and Barbarin, 2008). Depending on which effect prevails, a simple comparison of schools which have improved their infrastructure with those which have not might either underestimate or overestimate the real impact of such policy in student grades.

One good identification strategy in these cases is to control for the socioeconomic levels of each school in each year, in the econometric models. However, a yearly record of these data in the Portuguese case is not made available by the Ministry of Education, which has only made public the data for the academic year of 2012/2013 in its *Dados de Contexto das Unidades Orgânicas* report (DGE, n.d.c). Therefore, this identification strategy is not pursued in the present research (others will be). However, the data collected from this report will still be useful for some analyses performed later on.

The first socioeconomic indicators collected from the Ministry's report were the average educational attainment of the *fathers* of all students in a school, measured as the number of years of education received, and the average educational attainment of the *mothers* of all students in a school. These were compiled to provide a picture of the average parental educational levels in each school.

Household income is proxied by data collected on the percentage of students that, for belonging to underprivileged households, receive social support in the form of grants from the state to be used to finance their education. These grants have two main *ranks* – rank A refers to a larger grant that is given to students belonging to households whose gross earnings are below 2.934,54 euros/year, and rank B refers to a smaller grant that is given to students belonging to households with gross earnings between 2.934,54 euros/year and 5.869,08 euros/year (Segurança Social, n.d.) (the figures refer to the year of 2015, but these thresholds are subject to changes from year to year). Data for the percentage of students in each school who receives each of the two grants was compiled from the above-mentioned source.

Another variable that is likely to be correlated with the socioeconomic levels of the students in each school is the percentage of students that is enrolled in a regular *track* of secondary education. In the Portuguese secondary school system there are two main *tracks* that students can choose for their studies - a regular track, which is more scientifically-focused, and prepares students for proceeding to higher education studies in the future, and a technical track, with a more practical focus, tailored for students with lower previous academic performance who do not wish to proceed their studies at university level. Students who are not in regular *tracks* are more likely to come from poorer and less educated households, and therefore data for this indicator is collected in the Ministry of Education's report. Data on the rates of completion of secondary studies was also gathered from the same source, as non-completion and retention rates are correlated with the socioeconomic status.

5. Experimental Design

The aim of the present research is to assess the impact of the improvement in the physical and technological conditions of the Portuguese secondary schools involved in the SSMP, by evaluating the subsequent changes in educational outcomes of students belonging to *intervened schools*, relative to the outcomes of *not-intervened* schools (which includes *not-selected* and *on-hold* schools).

5.1 Threats to validity

In order to estimate the causal effect of the SSMP on academic outcomes of students, a simple comparison of academic outcomes between intervened (treated) and not-intervened (control) schools is insufficient. Two main threats to the validity of the results may arise.

First, we could potentially be running into selection problems. In concrete, if treated schools are fundamentally different from control schools in aspects that are correlated both with academic outcomes and selection to the program, the estimated impact of the program could be biased by these unobserved differences. For example, if treated schools happen to have higher levels of physical deterioration prior to the program, and this was part of the criteria for which they were chosen to be intervened, then we will be underestimating the impact of the SSMP. The same holds if treated schools happen to differ from control schools in their socioeconomic status.

Second, time-specific variation may be affecting the dependent variable. For example, if exams happen to be easier in the years following the implementation of the SSMP, leading to higher average grades, we may incorrectly interpret that the SSMP increased the academic preparation of students.

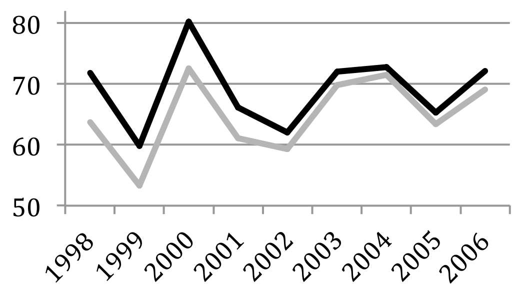
In what is concerning the first threat to validity, the threat of *selection bias*, it is worth it pointing out that no concrete criteria was used when selecting the schools to be intervened under the SSMP. In fact, the possibility of intervention under the SSMP program was given to all public schools in the country, regardless of their buildings' structural problems, socioeconomic contexts or locations, leaving the decision of signing up for the modernization process completely to the discretion of each school's management board. While this consideration may mitigate some of the risk of selection bias, it would be a failure of judgement to assume that selection to the SSMP was completely random. In fact, while the Ministry of education surveyed all schools to assess their willingness to participate in the program, these invitations came particularly sooner, and with a special focus, to older schools (with over 50 years of existence) and to schools located in *less developed regions,* as defined by the European Union's cohesion policy (specifically the *Norte, Centro* and *Alentejo* regions), as interventions in these regions were eligible for EU funding

(Parque Escolar, n.d.b). And even if we discard the Ministry's different approaches to different types of schools, it is only intuitive to think that schools showing larger signs of physical and functional degradation, as well as schools with higher levels of underprivileged students, were more likely to accept the Ministry's invitation to be part of the SSMP.

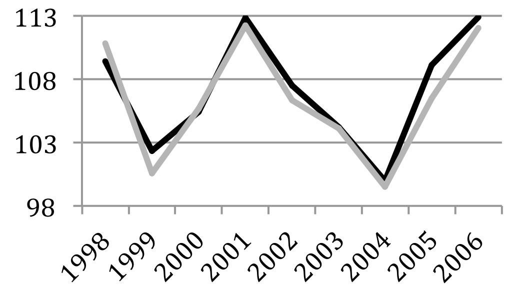
As a result, it is possible that these fundamental differences between treated and control schools have repercussions in academic achievement. In order to test this, for every year before the beginning of the implementation of the SSMP (2007), the average grades in each year, in each exam, for the pool of schools that was intervened under the SSMP are plotted in figure 1 (dark line), against the average grades of the schools that were not intervened (grey line). Also, an average for each group of schools on a number of socioeconomic indicators is computed and presented in table 5: the average education of fathers, the average education of mothers, the percentage of students receiving a grant of rank A and rank B, the percentage of students in regular tracks of secondary education and the percentage of students concluding their high school studies are shown below. As already mentioned, these data are from the academic year of 2012/2013. However, since the socioeconomic indicators of students are very closely correlated with the geographical location of schools, which does not change over time, these figures can be assumed to represent the socioeconomic situation of the schools in all the years of analysis.

From the data presented in the figures it is possible to make some interesting conclusions. The schools that were intervened under the SSMP seemed to be in better condition, academically and socioeconomically, than schools that were not intervened by the SSMP, before the start of the program. Starting with figure 1, it is clear that, in almost all years since 1998 and until the beginning of the SSMP in 2007, average grades in all of the evaluated exams were higher for schools that ended up being modernized, relative to schools that were not, the only exception being the Portuguese exams of 1998 and 2000, in which the average grade for control schools slightly surpassed that of treated schools. Table 5 also shows the advantage of treated schools over control schools. Treated schools have a lower percentage of their student populations receiving financial support from the state, either through rank A or rank B, and they have higher levels of parental education, both in

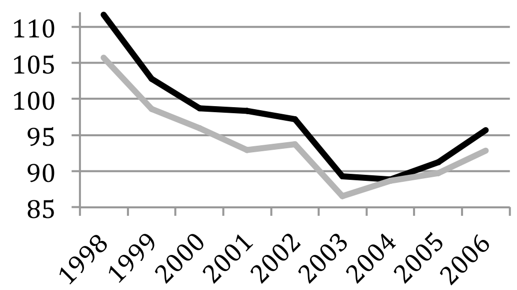
Figure 1: Pre-trends in average grades for selected and not-selected schools



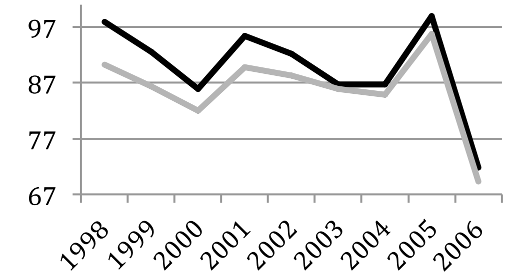
(a) math



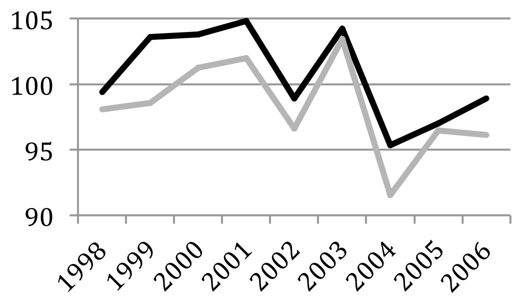
(b) portuguese



(c) biology



(d) physics



(e) rest

the case of fathers and mothers. The percentage of students in regular tracks of secondary education is also higher for treated schools. However, a slightly higher percentage of students in control schools concluded their high school studies.

Regardless of the reasons behind the differences between treated and control schools, the fact that these differences exist proves that the first threat to validity is a real threat to a consistent estimation of the impact of the SSMP. Since schools that

Table 5: Socioeconomic indicators in selected and not-selected schools

	Treated	Control
Schools	167	473
% students in regular tracks	72,32	71,93
% students social support (rank A)	13,46	17,3
% students social support (rank B)	14,33	17,72
Average Father's Education	9,61	8,45
Average Mother's Education	10,24	9,22
% Concluding Secondary	61,85	62,02

ended up being treated by the SSMP had both higher academic achievement and socioeconomic standards than the rest of the school prior to the intervention, unless these differences are accounted for the impact of the program is likely to be overestimated.

As for the second threat to validity, it is clear from figure 1 that there is significant variation from year to year in the average grades for all exams, and that this variation does not follow a specific time pattern or trend. While it is true that part of this variation could be attributed to different average academic skills of each cohort, there is no reason to believe that students from one cohort should be so significantly different from their preceding or succeeding cohorts in academic skills, so as to cause such variation in average grades. Assuming (plausibly) that the quality of education (i.e. quality of teachers, quality of teaching methods etc.) remains rather constant from year to year, the only possible explanation for this variation in average grades should be that exams differ in their difficulty from one year to the other. The starkest example of this situation occurred in 2008 with the math exam, when there was a larger-than-usual increase in the students' performance at the exam (not visible in figure 1) which was not due to an increase in that year's students academic skills relative to the previous year's cohort of students, but to a much easier-than-usual exam in that year, as reported by many students and media outlets across the country ("Provas de Matemática", 2008). Concluding, the threat of time-specific variation is a real threat to a consistent estimation of the impact of the SSMP.

5.2 Identification Strategy

Given the above-described threats, what is worthy of analysis is thus how *evolutions* in academic outcomes, and not their absolute values, compare between treated and control schools. As such, a difference-in-differences approach (DD-model) and a fixed-effects approach (FE-model) are taken in the present research.

5.2.1 Difference-in Differences Model

The purpose of the DD-model is to first look at the difference of outcomes pre and post treatment, for both treatment and control schools separately – the first differences - and then take the difference between those evolutions – the second difference. By performing the first difference, the confounding features of each group of schools, namely the schools' level of physical degradation and the schools' socioeconomic contexts, are being netted out. By performing the second difference, the time-specific variation that could be affecting the outcome is being netted out.

In table 6, the setup of the DD-model is presented, and the main outcome of interest is the last cell, (D-C)-(B-A), where B and D stand for the outcomes after (or during) the period of intervention for control and treated schools, respectively.

Table 6: Difference-in-Differences model

	Control Group: Not-intervened Schools	Treatment Group: Intervened Schools	Av. Grade Differences
Before	A	C	C - A
After/During	B	D	D - B
Change in Av. Grades	B - A	D - C	(D – C) - (B - A)

In order to compare treated with control schools in the appropriate time periods, in the DD model, one needs to determine the periods in which control schools would have been intervened, had they been chosen for the SSMP. This counterfactual is obviously impossible to observe and thus another methodological choice has to be made. For all control schools, the dates delimiting the time periods *during* and *after* were defined resorting to the average dates of beginning and ending of the modernization process for treated schools, and then using the 6-month rule as explained in section 4.3. For the treated schools, the average beginning date of the modernization process is June of 2009, and the average ending date of the modernization process and reopening of the school is February of 2012. Thus, from the above-described methodology comes that, for all control schools, the period *during* is defined as beginning in 2010 and ending in 2012 and, consequently, the period after was defined as beginning in 2013 and ending in 2014 – the last year of available data. Possible implications to the overall quality of the results coming from this methodological choice are discussed in the *Results* section.

This transformation resulted in a different sample from that of table 4, in which control schools now have observations for the time periods *during* and *after.* This sample is summarized in table 7.

Table 7: Sample of observations for the DD model, showing observations by status of intervention under the SSMP and time period

Treated Schools				
		Av. Grades		
	Before	During	After	TOTAL
Intervened	10 334	1 682	2 039	14 055
Control Schools				
		Av. Grades		
	Before	During	After	TOTAL
Not Selected	16 267	4 542	3 201	24 010
On-hold	7 029	1 800	1 211	10 040
			TOTAL	48 105

The regression for the DD-model will then be:

$$Grade_{jit} = \beta_0 + \beta_1 Program_{it} + \beta_2 During_{it} + \beta_3 After_{it} + \beta_4 Program_{it} * During_{it} + \beta_5 Program_{it} * After_{it} + \varepsilon_{it}$$

$Grade_{jit}$	represents, for each exam *j*, the average grade obtained by school *i* in year *t*.
$Program_{it}$	is a dummy variable equal to one for schools that were intervened under the SSMP and equal to zero otherwise.
$During_{it}$	is a dummy variable equal to one if a certain grade corresponds to an exam taken in a year during which modernization works were still under way in the respective school.
$After_{it}$	is a dummy variable equal to one if a certain grade corresponds to an exam taken in a year after the modernization works have been completed in the respective school.
i	represents the school.
t	represents the year.
ε_{it}	represents the error term.

In conclusion, in the difference-in-differences model, the impact of the SSMP post-modernization will be captured by β_4, and β_5 will capture the effect of the SSMP during the modernization phase.

5.2.1.1 Common-trend assumption

The main the identifying assumption in the DD-model is that, without the program, average grades in treated schools would have followed the same trend as that experienced by average grades of control schools after the program was implemented. This is what is usually referred to as the *common trend assumption* and it is crucial to ensure that time trends that could be affecting the outcome variable are effectively being netted out.

The plausibility of this assumption can be checked by comparing the pre-trend in average grades between treated and untreated schools – the development of grades prior to the kick-off of the SSMP in 2007 – as done in figure 1.

In all exams, academic outcomes on treated and control schools follow a similar pattern: a increase (decrease) in average grades, in a certain year, in treated schools, is always accompanied by an increase (decrease) in average grades in control schools, suggesting that the *common trend assumption* holds for these data, and that the difference-in-difference model can be used to consistently estimate the impact of the SSMP. This was an expected outcome since all time-related factors that affect the outcome variable do, in fact, affect all schools in the same way. For example, the level of difficulty of the nationwide exams in a certain year affected all schools without exception, as did the government's 2010 policies of increasing classroom sizes, reducing the number of classrooms and extending the mandatory education to 12 years (Capucho, 2012; Paula, 2013).

5.2.2. Fixed Effects Model

As an alternative to the DD approach, a new model will be created with fixed-effect regressors for each school, so as to capture all the confounding school-level variation in grades. Since socioeconomic indicators are assumed to be time-invariant, they are not included in the model. A control variable for the year is included to isolate the time-specific variation in average grades (due to exams being more/less difficult in a certain year). The sample of observations for this model is, thus, that of table 4.

The regression for the FE-model will then be:

$$Grade_{jit} = \Phi_i + \alpha_1 \, Program_{it} * During_{it} + \alpha_2 \, Program_{it} * After_{it} + \sum_{t=1998}^{2015} \delta_t T_t + \varepsilon_{it}$$

$Grade_{jit}$	represents, for each exam *j*, the average grade obtained by school *i* in year *t*.
Φ_i	is the unobserved individual effect.
$Program_{it}$	is a dummy variable equal to one for schools that were intervened under the SSMP and equal to zero otherwise.
T_t	represents the year in which the exam took place.
$During_{it}$	is a dummy variable equal to one if a certain grade corresponds to an exam taken in a year during which modernization works were still under way in the respective school.
$After_{it}$	is a dummy variable equal to one if a certain grade corresponds to an exam taken in a year after the modernization works have been completed in the respective school.
δ_t	is the coefficient on the dummy variable T_t
T_t	is a dummy variable equal to one for year t, and zero otherwise
i	represents the school.
t	represents the year.
ε_{it}	represents the error term.

α_1 will capture the effect of the SSMP during the period of intervention, and α_2 will capture the effect of the SSMP after the period of intervention.

6. Results

Before discussing the results of the models, it is worth pointing out that the estimated impact of the program in the period *during* is expected to be negative since the construction works are likely to create a harmful environment for learning, and the estimated impact of the program in the period *after* is expected to be positive because the modernized school infrastructure is expected to create a favourable environment for learning.

To gain a better understanding of the differences between treated and control schools, table 8 presents the summary statistics of the DD model for the average grades in each group of schools, and for each of the defined time periods.

In the period *before,* all exams except Portuguese have statistically significant differences in average grades between treated and control schools at a 1% level, reflecting the socioeconomic advantage of treated schools. In the period *during*, biology, math and physics have statistically significant differences in average grades between treated and control schools, at a 1%, 5% and 10% level, respectively, as indicated by the t tests. In the period *after*, Portuguese and rest have statistically significant differences in average grades between treated and control schools, at a 1% and 10% level, respectively. No inference can, however, be made from these results because these still reflect the confounding selection and time-specific variation effects that the DD model is expected to eliminate.

However, a simple across-group difference of the within-group differences in average grades between periods, as done in the last column of table 8, suggests that the expected negative impact of the program in the period *during* is visible for all the exams. As for the expected positive effect of the program in the period *after,* it is only visible for the *Rest* group, since in all the other courses the effect is negative.

Table 8: Summary Statistics of the DD model

	Treated	Control	p	Difference Treated	Difference Control	Diff in Diff
Schools	167	473				
Observations	14 063	33 899				
			Math			
Av. Grade Before	77,24	73,94	***			
Av. Grade During	85,48	89,46	**	8,24	15,52	-7,28
Av. Grade After	79,64	79,81		2,4	5,87	-3,47
			Portuguese			
Av. Grade Before	105,84	105,86				
Av. Grade During	94,97	95,93		-10,87	-9,93	-0,94
Av. Grade After	94,51	98,18	***	-11,33	-7,68	-3,65
			Biology			
Av. Grade Before	96,99	94,41	***			
Av. Grade During	93,8	97,43	***	-3,19	3,02	-6,21
Av. Grade After	91,08	89,64		-5,91	-4,77	-1,14
			Physics			
Av. Grade Before	88,67	85,21	***			
Av. Grade During	85,01	87,1	*	-3,66	1,89	-5,55
Av. Grade After	80,26	80,57		-8,41	-4,64	-3,77
			Rest			
Av. Grade Before	100,88	99,66	***			
Av. Grade During	96,03	96,05		-4,85	-3,61	-1,24
Av. Grade After	98,04	96,25	*	-2,84	-3,41	0,57
			Socioeconomic Indicators			
Rank A	13,46	17,30				
Rank B	14,33	17,72				
Father Educ	9,61	8,45				
Mother Educ	10,24	9,22				
Regular Track	72,32	71,93				
Concluding	61,85	62,02				

Table 9 presents the summary statistics of the FE model for the average grades in each group of schools. The difference in average grades between treated and control schools for the biology, physics and rest exams are statistically significant at a 1% level, while math is only significant at a 10% level. The difference in average grades between treated and control schools for the Portuguese exam does not seem to be statistically significant. Again, no inference can be made from these results because these still reflect the confounding selection and time-specific effects that the FE model is expected to eliminate.

Table 9: Summary Statistics of the FE model

	Treated	Control	Diff.	
Schools	167	473		
Observations	14 063	33 899		
Av. Grade Math	78,57	77,56	1,01	*
Av. Grade Portuguese	102,9	103,04	-0,14	
Av. Grade Biology	95,75	94,35	1,40	***
Av. Grade Physics	87,01	84,95	2,06	***
Av. Grade Rest	99,88	98,55	1,33	***
Socioeconomic Indicators				
Rank A	13,46	17,30		
Rank B	14,33	17,72		
Father Educ	9,61	8,45		
Mother Educ	10,24	9,22		
Regular Track	72,32	71,93		
Concluding	61,85	62,02		

Below, the estimation results for both the FE and the DD model are discussed. The outcome variable – the average exam scores in each school – was standardized to have a mean of zero and a standard deviation of one. This procedure is common in the educational policy literature as it allows for an easier interpretation of the results: each estimated value for the standardized variable indicates its difference from the mean of the original variable, in number of standard deviations (of the original variable). The use of this strategy in this paper also allows for comparisons of the magnitude of the estimates with other papers on education policy that have followed the same standardization approach, some of the most widely cited being Duflo (2000) and Krueger (1999). It also enables to put some perspective on the size of the estimated effects. Cohen (1998) defined a standardized effect of 0.2 standard deviations as a "small" effect, a standardized effect of 0.5 standard deviations as a "medium" effect and a standardized effect of 0.8 standard deviations as a "large" effect.

6.1 Difference-in-Differences Model

The estimation results of the DD-model are presented in table 10. The coefficient of the interaction between the variables *program* and *during* – in the model specified as *progdur* - is expected to be negative, and the coefficient of the interaction between the variable *program* and *after* – in the model specified as *progaft* - is expected to be positive.

Table 10: Main DD estimates of the effect of the program on student performance

	Math	Portuguese	Biology	Physics	Rest
	(1)	(2)	(3)	(4)	(5)
Program	.141***	-.002	.159***	.190***	.071***
	(.026)	(.025)	(.026)	(.026)	(.026)
During	.664***	-.693***	.187***	.104***	-.208***
	(.031)	(.030)	(.031)	(.031)	(.032)
After	.251***	-.536***	-.295***	-.256***	-.197***
	(.036)	(.035)	(.036)	(.036)	(.037)
progdur	-.312***	-.065	-.383***	-.305***	-.072
	(.065)	(.063)	.066	(.066)	(.066)
progaft	-.149**	-.254***	-.071	-.207***	.033
	(.064)	(.062)	(.065)	(.065)	(.065)
Observations	9661	9709	9587	9607	9541
R-squared	.051	.096	.020	.018	.011

The results in table 10 suggest different conclusions for the science and engineering (S&E) related exams of math, biology and physics than for those exams that are not science-related (the Portuguese exam and the five exams that constitute the Rest average grade). For the first group of exams, *progdur* is, as expected, negative and statistically significant at a 1% level. The magnitude of this impact is estimated to range between -0,305 standard deviations in the case of the physics exam and -0,383 standard deviations in the case of biology, thus representing "small" size effects as defined by Cohen (1998). For the second group of exams, estimates are also negative but not statistically significant.

In what is concerning *progaft*, the estimates obtained for all exams except rest are negative, contrary to what was expected, but only statistically significant at a 1% confidence level for the case of Portuguese and physics, and at a 5% level for math.

These estimated effects range between -0,149 standard deviations in the case of math and -0,254 standard deviations in the case of Portuguese, again “small” effects.

These results should be interpreted having in mind that the difference-in-differences model was constructed upon two methodological choices. The first choice was the definition of the time period *during* as beginning in a certain year only if construction in the school started at least six months prior to the exam of that year, and of the time period *after* as beginning in a certain year only if the construction in the school has ended at least six months prior to the exam of that year. This has the potential for decreasing the overall quality of the results if the effects of the construction works, in the case of the period *during,* and the effects of the improved infrastructure, in the case of the period *after,* take longer/less time than six months to create a measurable change in the exam grades. However, this drawback is addressed in the *Sensitivity Analysis* section of this paper, in which the time periods *during* and *after* are defined using both longer and shorter time delays. The second methodological choice concerns the definition of the time periods *during* and *after* for control schools resorting to the average beginning and ending dates of the modernization processes in treated schools. This is the best approximation possible but it is still arbitrary as one cannot tell when these schools would have been intervened had they been chosen for the SSMP, since this outcome is not observable. Thus, despite the fact that the underlying problem of defining time periods for control schools cannot be addressed in a better way, the methodological choice used in doing so can still be unsuited for the data and it cannot be challenged in the sensitivity analysis. That is why the fixed-effects model is likely to render more accurate results than the difference-in-differences model.

6.2 Fixed Effects Model

The estimation results of the FE-model are presented in table 11. Again, the coefficient of the interaction between the variables *program* and *during* is expected to be negative, and the coefficient of the interaction between the variable *program* and *after* is expected to be positive.

Table 11: Main FE estimates of the effect of the program on student performance

	Math	Portuguese	Biology	Physics	Rest
	(1)	(2)	(3)	(4)	(5)
progdur	-.198***	-.131***	-.145***	-.139***	-.086*
	(.030)	(.042)	(.037)	(.036)	(.046)
progaft	-.283***	-.235***	-.231***	-.249***	-.112**
	(.030)	(.042)	(.036)	(.036)	(.046)
Observations	9661	9709	9587	9607	9541
R-squared	.646	.204	.415	.376	.171

The results in table 11 suggest that, for all exams, the estimates for *progdur* are negative, as expected. Estimates are statistically significant at a 1% level for all exams except Rest, whose estimate is only significant at a 10% level. Estimates for S&E exams are more negative than for Portuguese and Rest. The estimates range from -0,086 standard deviations in the case of Rest and -0,198 standard deviations in the case of math, thus “small” effects in all cases.

The estimates for *progaft* are negative for all exams, contrary to the expected, and statistically significant at a 1% confidence level for all exams except rest, which is only significant at a 5% level. Estimates range from -0,112 standard deviations in the case of rest to 0,283 standard deviations in the case of math, again, “small” effects.

7. Sensitivity Analysis

The use of a more accurate FE model already allows to test the robustness of the less accurate DD estimates. However, in order to test the robustness of the results of section 6 even further, two other sensitivity analyses are performed.

First, the definition of the time periods *during* and *after* are changed, to accommodate a shorter delay and a longer delay in obtaining measurable effects on exam grades coming from the school status (before construction, during construction or after construction). In this way, instead of the 6 months of delay that were defined in section 4.3 as the minimum amount of time needed for a change in the construction status of a school to produce measurable changes in grades of the following exam, I allow for a shorter period of 3 months of delay and a longer period of 9 months of delay. Since the academic year starts in September and ends in June, a 12-month delay would produce almost the same results as a 9 months delay, since the additional three months would fall in the summer months, in which schools are not functioning and thus not producing an impact on the preparation of students for the exams. Thus, a 12-month delay analysis will not be included.

Second, to test the robustness of the results even further, two new samples were created, to which the regressions specified in section 5.2 are applied. In the first new sample, observations from private schools were excluded. These schools were all included in the control group since the program was directed at public schools only. The rationale behind their exclusion from the sample is that there is no information on whether some of these schools have gone through their own privately-financed modernization processes (outside the scope of the SSMP), which is likely that some of them have. If this is the case, by including private schools in the sample we are incorrectly including the effect of these modernizations in the overall effect of the SSMP, and thus we can be underestimating the impact of the SSMP in section 6. Therefore, the estimates obtained from this new sample will be more accurate. In the second new sample, the *not-selected* schools (schools that were never selected to be part of the SMMP) are excluded from the sample, leaving only *on-hold* schools in the control group. These *on-hold* schools are schools that were selected to be intervened by the SSMP but ended up not being intervened due to financial constraints. If there are unobserved differences in socioeconomic and structural characteristics between treated and control schools, leaving in the control group only schools that were supposed to receive treatment can more effectively eliminate these differences and make treatment and control groups more comparable, in this way further reducing the risk of selection bias. For this reason, the use of this sample will produce the most reliable estimates.

7.1 Difference-in-Differences Estimates

Table 13 of the appendix presents the same results of the DD model described in table 10, together with the results obtained from the use of the 3-month and 9-month definitions of the time variables *during* and *after.* Table 14 of the appendix presents the same estimates but for the new sample in which private schools were excluded, thus improving the quality of the control group. Table 15 of the appendix presents the same estimates but for the sample in which only *on-hold* schools were included in the control group, and in this way further improving the quality of the control group.

For *progdur,* no significant changes were registered in the estimates of all the exams across all the specifications used. The estimates for math, biology and physics remain negative and statistically significant at a 1% level regardless of the sample used or the definition of time periods used. The estimates for a 9-month definition of time period are always more negative, some of which even pointing to a "medium" size negative effect. The estimates for the Portuguese exam, which become statistically significant at a 5% level in the initial sample (table 13) when using a 3 and 9-month definition for time periods, loose all significance when using the two preferred samples (tables 14 and 15), regardless of the definition of the time periods used. The estimates for rest are always not statistically significant, regardless of the samples and definitions of time periods used.

For *progaft*, the definition of time period and the sample used causes estimates to change substantially in size, sign and significance. In the initial sample, the estimates for math, Portuguese and physics loose all significance when a 9-month definition for time periods is used (table 13). When excluding public schools from the sample (table 14), a set of changes occur in the estimates: the estimates for math become significant only at a 10% level, the estimate for biology with a 9 month definition of time periods becomes statistically significant at a 5% level, the estimates for physics loose all their significance and the estimates for Rest, which are positive, gain statistical significance at a 10% and 5% level, depending on the definition of time periods used. When using only *on-hold* schools in the control group, the preferred sample (table 15), the estimates for math, biology and physics are only statistically significant using a 9-month definition of time periods and are positive. Estimates for rest are positive and statistically significant at a 10% or 5% level depending on the definition of time periods used and the estimates for Portuguese loose all statistical significance. In all cases, the statistically significant effects estimated are of a "small" size, as defined by Cohen (1988).

7.2 Fixed-Effects Estimates

The FE model provides more reliable estimates than the DD model, for the reasons described in section 6.1. Table 16 of the appendix presents the same results of the FE model described in table 11, together with the results obtained from the use of the 3-month and 9-month definitions of the time variables *during* and *after.* Table 17 of the appendix presents the same estimates but for the new sample in which private schools were excluded, thus improving the quality of the control group. Table 18 of the appendix presents the same estimates but for the sample in which only *on-hold* schools were included in the control group, in this way increasing the quality of the control group even further.

The estimates for *progdur* vary across specifications in size and significance. Despite the fact that, in the main sample (table 16), the different definitions of time periods did not alter significantly the estimates obtained, when private schools are excluded from the sample (table 17) the estimates for Portuguese and Rest loose all their significance, and the estimates for math, biology and physics seem to decrease in size (become less negative). When only *on-hold* schools are included in the control group, the preferred sample (table 18), the estimates for physics also loose their statistical significance except for the 9-month estimate which is statistically significant at a 10% level. The 6-month estimate for biology also becomes significant only at a 10% level. In all cases, the estimated effects are of a “small” size, as defined by Cohen (1988).

In the case of *progaft,* the different definitions of time periods also did not alter significantly the estimates obtained in the main sample (table 16). When excluding public schools from the sample (table 17), only the estimates for Rest seem to change by loosing their statistical significance (only the 3-month estimate is significant at a 10% level). When using only *on-hold* schools in the control group, the preferred sample (table 18), the estimates for Portuguese and physics become less statistically significant and smaller (less negative). Again, all estimated effects are of a “small” size as defined by Cohen (1988).

8. Discussion

Overall, the results seem to point to two main findings. First, the SSMP seems to have caused a small to medium size decrease in academic performance of students during the period in which constructions works were occurring (here defined as the period *during)* and this impact was especially significant in the case of S&E exams: math, biology and physics. This is finding is consistent with the literature and with the facts on the ground, since schools have seen their S&E-specific infrastructure (such as labs) demolished and not rebuilt for some years. The size of this impact ranges between -0,126 and -0,563 standard deviations in the several DD estimates obtained for the different exams and between -0,053 and -0,214 in the more reliable FE estimates. Second, although some DD estimates point to a possible positive impact of the SSMP in academic performance of students in the period *after* construction works have ceased, the more credible FE estimates suggest that the program actually continues to negatively affect student outcomes in all exams (these impacts are, again, especially strong and statistically significant for S&E exams). In most cases, this fall in academic performance is even stronger than the one occurred in the period *during*. These results are contrary to the predictions made, as the aim of the policy was to increase educational standards of students.

It is quite counterintuitive that a program such as the SSMP, which aimed at improving the physical conditions of schools, ended up having a negative impact on student outcomes. One possible explanation is that the changes operated in schools by the SSMP were not the most adequate ones and, as a result, they have decreased the quality of the learning environments to levels that were even lower than those prior to the SSMP. In fact, a number of different complaints from the schools which have participated in the SSMP were subject of nationwide media attention: from schools that re-opened without all the facilities that were projected to be built to non-functioning air conditioning systems, many examples of unsatisfied school directors, staff and students became public (Araújo, 2011). While some of these complaints may be factual and reasonable, it is hard to believe that their negative effect outweighs all the positive transformations that the SSMP has made possible, especially when the OECD itself evaluated the program as having created “significant achievements” through the “radically improved” quality of the buildings.

Another, perhaps more plausible, explanation for such results stems from the fact that education is a continuous and cumulative learning process and, thus, if a certain student does an exam in a modernized school at the end of her high school education, her performance can still be lower that that of students in non-modernized schools if her school was under construction in the first years of her high school education. In other words, if a student happens to experience both the *during* and the *after* periods during her high school years, the negative impact on her outcomes created in the *during* period can create a disadvantage that still has repercussions on her later outcomes, even when the school is already modernized and has re-opened to its normal functioning.

It is fairly easy to test whether these cross-year negative spillovers actually took place. High school education in Portugal is comprised in three years of education (10^{th} to 12^{th} grade). The Portuguese and Math exams are conducted at the end of the third year, while all the other exams are conducted at the end of the second year. Thus, it is easy to divide all observations in the sample belonging to the period *after* in two different groups: first, the average grades in years in which students that made the exam had a full 3 years of high school education in a modernized school (and thus were never affected by the construction works); second, the average grades in years in which the students that made the exam had at least one year (but no more of two) of their studies in a school that was under construction.

Table 12 shows the evolution of average grades in the period *after* of these two groups between 2012 and 2014, since these are the only years in which there are enough observations to make comparisons possible.

As table 12 shows, it is not for all cases that, in each year, students who spent all three years of their high school education in a modernized school do better than those who have spent at least one year in a school under construction. This is only the case math in 2012 and 2014 and for biology, physics and rest in 2013 and 2014. Thus the negative year-to-year spillover argument does not by itself explain all the negative effects of the SSMP. However one should note that these differences do not take into account possible disparities in socioeconomic status of the schools in each group and that they are only statistically significant in three circumstances: math in 2012 and 2014 (positive) and Portuguese in 2012 (negative). This is maybe due to the relatively low amount of observations that one is able to collect for each of the two groups given that the program is still very recent and there are not many schools already in the period *after* the construction. Therefore, no clear conclusions should be taken from this exercise until we have enough data to make comparisons of the two groups credible – that means, we will have to wait a few more years to see more schools completing their modernization and to have more cohorts studying in modernized schools for their entire three years of high school.

Table 12: Evolution of average grades in the period *after* intervention

		Av. Standardized Grades								
		2012			2013			2014		
	Exams of students who	Obs.	Av.	p	Obs.	Av.	p	Obs.	Av.	p
Math	Spent all 3 years in a modernized school	18	0,25		52	-0,01		31	-0,06	
	Spent at least 1 year in a school under construction	77	-0,11		64	0,11		95	-0,27	
	Diff		0,36	**		-0,13			0,22	*
Portuguese	Spent all 3 years in a modernized school	18	-0,88		52	-1,20		31	-0,01	
	Spent at least 1 year in a school under construction	77	-0,51		64	-0,98		95	0,12	
	Diff		-0,37	**		-0,22			-0,13	
Biology	Spent all 3 years in a modernized school	52	-0,62		95	-0,93		116	0,23	
	Spent at least 1 year in a school under construction	43	-0,45		21	-1,07		10	0,09	
	Diff		-0,16			0,14			0,14	
Physics	Spent all 3 years in a modernized school	52	-0,79		95	-0,40		116	-0,19	
	Spent at least 1 year in a school under construction	43	-0,69		21	-0,56		10	-0,45	
	Diff		-0,10			0,16			0,26	
Rest	Spent all 3 years in a modernized school	52	0,13		95	-0,15		116	0,01	
	Spent at least 1 year in a school under construction	43	0,17		21	-0,19		10	-0,17	
	Diff		-0,04			0,03			0,18	

9. Conclusion

This study exploits the variation in spending on school infrastructure coming from the SSMP – a secondary school modernization program conducted in Portugal – in order to estimate the impact of the modernization of school buildings on student achievement. Using a difference-in-differences and fixed effects approach, the impact of this program in the evolution of student grades in a set of exams was estimated, by comparing outcomes between treated and control schools.

The results of this research point to a small to medium-size negative impact of the program on student grades during the phase of the program in which renovation works are occurring in the schools. This result is in line with the experiences of several intervened schools which have reported a significant worsening of the learning environment during this period, due to overly high levels of noise and air pollution and the displacement of students less-equipped modular classrooms. The reduction in student performance is especially significant in the S&E related courses of math, biology, geology, physics and chemistry, as S&E-specific infrastructure such as science labs are frequently demolished, and the new prefabricated classrooms do not offer the necessary conditions for a satisfactory preparation for the exams of these courses. These were also the courses for which the estimates proved to be more robust to all the models and specifications applied in the sensitivity analysis.

The more credible FE estimates also point to a small, but statistically significant, negative impact of the program in the years shortly after the renovation works took place and the modernized schools re-opened to their normal academic functioning. Again, the estimates for the S&E exams are the most statistically significant and robust to the all the models and specifications applied in the sensitivity analysis. These results are contrary to the positive impact of the modernized schools in students' outcomes predicted by the OECD's evaluation of the SSMP, as well as the large bulk of the literature on this topic.

More research at a further date is needed to make sense of these results. As the program is very recent, and there are only a few schools that have completed their modernization process, data on academic outcomes for more years will tell whether negative year-to-year spillovers are the cause of the negative estimates. If this proves to be the case, it is reasonable to say that the program created a disadvantage only for those students who started their studies in a school that was experiencing construction works but did their exams after these works were finished. In contrast, all the students who go through their full high-school years in a modernized school will see their academic outcomes increase as a result of the SSMP.

If, on the contrary, negative year-to-year spillovers are ruled out, more years of data on academic outcomes will also tell whether what appears to be the negative impact of the program after its completion is a short-term or a long-term one. If it is short-lived, it could be that there is period of adaptation of the students and school-staff to the new infrastructure that creates a temporary fall in grades, but ends up turning into a long-term period of enhanced academic outcomes. If, on the contrary, the negative trend persists, the only reasonable explanation is that the changes in schools operated by the SSMP were not adequate and were in fact very harmful, as they have decreased the quality of the learning environment to levels that are even worse than those prior to the SSMP.

References

Ainda há 90 escolas em obras no início do segundo period (2010, January 5). *Diário de Notícias*. Retrieved from: http://www.dn.pt/portugal/interior/ainda-ha-90-escolas-em-obras-no-inicio-do-segundo-periodo-1461916.html

Angrist, J.D., Lavy, V. (2002). New Evidence on Classroom Computers and Pupil Learning. *Economic Journal*, 112:482, 735-765.

Araújo, A. (2011, January 14). Parque Escolar entrega edifício que ainda está a meio das obras. *Diário de Notícias.* Retrieved from: http://www.dn.pt/portugal/centro/interior/parque-escolar-entrega-edificio-que-ainda-esta-a-meio-das-obras-1756940.html

Bowers, JH and Burkett, GW (1987). The Relationship of Student Achievement and Characteristics in Two Selected School Facility Environmental Settings. 64th Conference of the Council for Educational Facility Planning International (CEFPI), Alberta, Canada.

Buckley et al. (2014). The Effects of School Facility Quality on Teacher Retention in Urban School Districts. *National Clearinghouse for Educational Facilities*. Retrieved from: http://www.ncef.org/pubs/teacherretention.pdf

Capucho, J. (2012, August 2). Escolaridade obrigatória publicada em Diário da República. *Diário de Notícas.* Retrieved from: http://www.dn.pt/portugal/interior/escolaridade-obrigatoria-publicada-em-diario-da-republica-2700525.html

Card, D. Payne, A. A. (2002). School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores. *Journal of Public Economics*, 83, 49-82.

Cellini, S.R., Ferreira, F., Rothstein, J. (2010). The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design. *Quarterly Journal of Economics*, 125, 215-261.

Chan, T. C. (1979). The impact of school building age on the achievement of eighth grade pupils from the public schools in the state of Georgia. Unpublished doctoral dissertation, University of Georgia, Athens.

Chay, K.Y., McEwan P.J., Urquiola, M. (2002). The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools. *Journal of Public Economics*, 83, 49-82.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, S; Evans, G; Krantz, DS; & Stokols, D. (1986). Behaviour, Health and Environmental Stress. New York, Plenum.

DETYA (Feb, 1993). In the Middle: Schooling for Young Adolescents (Compulsory Years of Schooling Project Paper No 7), Canberra, AGPS.

DGE (n.d.a). *Juri Nacional de Exames: Estatísticas (2004 – 2015)*. Retreived from: http://www.dge.mec.pt/estatisticas

DGE (n.d.b). *Juri Nacional de Exames: Estatísticas (Aqruivo 1998 – 2013)*. Retreived from: http://old.dge.mec.pt/jurinacionalexames/index.php?s=directorio&pid=33&ppid=4

DGE (n.d.c). Ministério da Educação e Ciência: Dados de Contexto das Unidades Orgânicas. Retrieved from: http://area.dge.mec.pt/jne2014/

Duflo, E. (2000). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. Working Paper 7860, National Bureau of Economic Research, August.

Earthman, G. (1998). The Impact of School Building Condition on Student Achievement and Behaviour. The Appraisal of Educational Investment Conference, Luxembourg, European Investment Bank and the Organisation for Economic Cooperation and Development.

Fisher, K (2001). Building Better Outcomes – The Impact of School Infrastructure on Student Outcomes and Behaviour. *DETYA.*

Frazier, L.M. (1993). Deteriorating school facilities and student learning. Washington D.C.: Office of Educational Research and Improvement.

Goolsbee, A., Guryan, J. (2006). The Impact of Internet Subsidies in Public Schools. *Review of Economics and Statistics*, 88, 336-347.

Guryan, J. (2002). Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts. NBER working paper no. 8269.

Inácio, A. (2014, January 13). Vinte e uma secundárias com aulas em contentores. *Jornal de Notícias.* Retrieved from: http://www.jn.pt/paginainicial/nacional/interior.aspx?content_id=3628721

Inspecção Geral das Finanças (2011). *Auditoria à Empresa Parque Escolar, EPE.* Retrieved from: http://www.parque-escolar.pt/docs/site/pt/programa/relatorio-igf.pdf

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics*, vol. 114, pp. 497-532.

Leuven, E., Lindahl, M., Oosterbeek H., Webbink, H.D. (2007). The Effect of Extra Funding for Disadvantaged Pupils on Achievement. *Review of Economics and Statistics* 89, 721-736.

Machin, S., McNally, S., Silva, O. (2007). New Technology in Schools: Is There a Payoff. *The Economic Journal*, 117(2007), 1145-1167.

McGuffey, C.W. And Brown, C.L. (1978). The impact of school building age on school achievement in Georgia. *CEFPI Journal*, 16, 6-9.

OECD (2009). *OECD/CELE Review of the Secondary School Modernisation Programme in Portugal*; Co-Authors: Rodolfo Almeida; Alastair Blyth David Forrester; Ann Gorey; Gaby Hostens.

OECD (2010). *The High Cost of Low Educational Performance*. Retrieved from: http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/thehighcostoflo weducationalperformance.htm

Papke, L.E. (2005). The effects of Spending on Test Pass Rates: Evidence from Michigan. *Journal of Public Economics*, 89, 821-839.

Parque Escolar (2011). *Parque Escolar conclui intervenções em 21 escolas e lança a Fase 4.* Retrieved from: http://parque-escolar.pt/pt/empresa/comunicados-imprensa-ver.aspx?id=48

Parque Escolar (n.d.a). *Historical Context.* Retrieved from: http://www.parque-escolar.pt/en/program/historical-context.aspx

Parque Escolar (n.d.b). *Strategic Context.* Retrieved from: http://www.parque-escolar.pt/en/program/strategic-context.aspx

Parque Escolar (n.d.c). *Tipificação das Intervenções.* Retrieved from: http://www.parque-escolar.pt/pt/programa/tipificacao-das-intervencoes.aspx

Parque Escolar (n.d.d). *Program Implementation Structure.* Retrieved from: http://www.parque-escolar.pt/en/program/programme-implementation-structure.aspx

Parque Escolar (n.d.e). *Programação e Faseamento do Investimento.* Retrieved from: http://www.parque-escolar.pt/pt/programa/faseamento-do-investimento.aspx

Parque Escolar (n.d.f). *Schools.* Retrieved from: http://www.parque-escolar.pt/en/schools/schools.aspx

Paula, H. (2013, June 13). Governo compromete-se a aumentar o número de alunos por turma. *Jornal de Negócios.* Retrieved from: http://www.jornaldenegocios.pt/economia/educacao/detalhe/fmi_recomenda_aumentar_o_numero_de_alunos_por_turma.html

Plumley, J. (1978). The Impact of School Building Age on the Academic Achievement of Selected Fourth Grade Pupils in the State of Georgia. D.Ed. Dissertation, Athens, Georgia, University of Georgia.

Provas de Matemática A e B foram "excessivamente fáceis" (2008, June 23). *Expresso.* Retrieved from: http://expresso.sapo.pt/dossies/dossiest_actualidade/ExamesNacionais20072008/provas-de-matematica-a-e-b-foram-excessivamente-faceis=f351366

Segurança Social (n.d.). *Abono de família para crianças e jovens.* Retrieved from: http://www4.seg-social.pt/abono-de-familia-para-criancas-e-jovens

Sinofsky, E. and Knirck, FG (1981). Choose the Right Colour for Your Learning Style. Instructional Innovator 26(3): 17–19.

Shor, I. (1996). When Students Have Power: Negotiating Authority in a Critical Pedagogy, Chicago and London: University of Chicago Press.

Tanner, K. and Jago, E. (1999). The Influence of the School Facility on Student Achievement. Washington D.C., University of Georgia.

Thomas, J.A. (1962). Efficiency in education: A study of the relationship between selected inputs and mean test scores in a sample of senior high school. Unpublished doctoral dissertation, Stanford University.

Wilkin, A., Kinder K., White, R., Atkinson, M. and Doherty, P. (2002). Research Summary: Towards the Development of Extended Schools, United Kingdom Department for Education and Skills and the United Kingdom National Union of Teachers.

Appendix

Table 13: DD estimates of the impact of the program on student performance, using and 6, 3 and 9 month definition of time periods.

Exam	Math			Portuguese			Biology			Physics			Rest		
Delay (months)	6	3	9	6	3	9	6	3	9	6	3	9	6	3	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Program	.141***	.138***	.140***	-.002	.022	-.011	.159***	.156***	.163***	.190***	.193***	.188***	.071***	.074***	.075***
	(.026)	(.026)	(.026)	(.025)	(.025)	(.025)	(.026)	(.026)	(.026)	(.026)	(.026)	(.026)	(.026)	(.027)	(.026)
During	.664***	.664***	.644***	-.693***	-.693***	-.657***	.187***	.187***	.202***	.104***	.104***	.116***	-.208***	-.208***	-.194***
	(.031)	(.031)	(.031)	(.030)	(.030)	(.030)	(.031)	(.031)	(.031)	(.031)	(.031)	(.031)	(.032)	(.032)	(.031)
After	.251***	.251***	.205***	-.536***	-.536***	-.498***	-.295***	-.295***	-.317***	-.256***	-.256***	-.290***	-.197***	-.197***	-.177***
	(.036)	(.036)	(.034)	(.035)	(.035)	(.033)	(.036)	(.036)	(.035)	(.036)	(.036)	(.035)	(.037)	(.037)	(.035)
Progdur	-.312***	-.330***	-.460***	-.065	-.126**	-.136**	-.383***	-.372***	-.563***	-.305***	-.311***	-.514***	-.072	-.082	-.092
	(.065)	(.064)	(.056)	(.063)	(.062)	(.054)	.066	(.065)	(.056)	(.066)	(.065)	(.056)	(.066)	(.066)	(.057)
Progaft	-.149**	-.133**	.059	-.254***	-.288***	-.093	-.071	-.027	.110	-.207***	-.182***	.019	.033	.040	.149
	(.064)	(.063)	(.089)	(.062)	(.061)	(.087)	(.065)	(.064)	(.090)	(.065)	(.064)	(.090)	(.065)	(.064)	(.091)
Observations	9661	9661	9661	9709	9709	9709	9587	9587	9587	9607	9607	9607	9541	9541	9541
R-squared	.051	.050	.048	.096	.101	.092	.020	.019	.022	.018	.018	.020	.011	.011	.011

Table 14: DD estimates of the impact of the program on student performance, using the sample without private schools.

Exam	Math			Portuguese			Biology			Physics			Rest		
Delay (months)	6	3	9	6	3	9	6	3	9	6	3	9	6	3	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Program	.284***	.282***	.286***	.116***	.140***	.116***	.327***	.324***	.335***	.332***	.334***	.334***	.210***	.213***	.221***
	(.024)	(.024)	(.024)	(.023)	(.023)	(.023)	(.024)	(.024)	(.024)	(.024)	(.024)	(.024)	(.024)	(.024)	(.024)
During	.597***	.597***	.577***	-.817***	-.817***	-.755***	.128***	.128***	.160***	.011	.011	.041	-.305***	-.305***	-.279***
	(.032)	(.032)	(.031)	(.030)	(.030)	(.030)	(.032)	(.032)	(.031)	(.032)	(.032)	(.031)	(.032)	(.032)	(.031)
After	.218***	.218***	.187***	-.677***	-.677***	-.582***	-.379***	-.379***	-.371***	-.374***	-.374***	-.377***	-.273***	-.273***	-.208***
	(.037)	(.037)	(.035)	(.036)	(.035)	(.034)	(.037)	(.037)	(.035)	(.037)	(.037)	(.035)	(.038)	(.038)	(.035)
Progdur	-.244***	-.263***	-.393***	.059	-.002	-.038	-.325***	-.314***	-.520***	-.212***	-.218***	-.439***	.025	.015	-.008
	(.060)	(.059)	(.051)	(.058)	(.057)	(.050)	(.060)	(.059)	(.052)	(.060)	(.059)	(.052)	(.060)	(.059)	(.052)
Progaft	-.115*	-.099*	.0769	-.113**	-.147***	-.010	.014	.057	.164**	-.089	-.064	.107	.110*	.116**	.181**
	(.059)	(.058)	(.080)	(.058)	(.056)	(.078)	(.060)	(.059)	(.081)	(.060)	(.059)	(.081)	(.060)	(.059)	(.081)
Observations	7947	7947	7947	7975	7975	7975	7909	7909	7909	7904	7904	7904	7895	7895	7895
R-squared	.061	.061	.058	.155	.163	.144	.047	.046	.050	.049	.049	.051	.037	.037	.036

Table 15: DD estimates of the impact of the program on student performance, using sample with only *on-hold* schools in the control group.

Exam	Math			Portuguese			Biology			Physics			Rest		
Delay (months)	6	3	9	6	3	9	6	3	9	6	3	9	6	3	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Program	.158***	.157***	.161***	.015	.039	.016	.226***	.222***	.234***	.180***	.183***	.182***	.089***	.091***	.099***
	(.028)	(.029)	(.028)	(.026)	(.026)	(.026)	(.029)	(.029)	(.030)	(.029)	(.029)	(.029)	(.028)	(.028)	(.027)
During	.559***	.558***	.545***	-.848***	-.848***	-.768***	.106**	.106**	.143***	-.050	-.050	-.010	-.297***	-.297***	-.268***
	(.049)	(.049)	(.048)	(.045)	(.045)	(.045)	(.050)	(.050)	(.049)	(.049)	(.049)	(.049)	(.047)	(.047)	(.047)
After	.132**	.132**	.105*	-.791***	-.791***	-.681***	-.387***	-.387***	-.376***	-.427***	-.427***	-.433***	-.280***	-.280***	-.214***
	(.057)	(.057)	(.054)	(.053)	(.053)	(.050)	(.059)	(.059)	(.055)	(.058)	(.058)	(.054)	(.056)	(.056)	(.052)
Progdur	-.206***	-.224***	-.361***	.091	.029	-.024	-.303***	-.292***	-.504***	-.151**	-.157**	-.388***	.018	.007	-.018
	(.069)	(.068)	(.062)	(.064)	(.063)	(.058)	(.070)	(.070)	(.063)	(.070)	(.069)	(.063)	(.067)	(.066)	(.060)
Progaft	-.029	-.013	.159*	.001	-.033	.090	.021	.065	.169**	-.036	-.011	.163*	.117*	.123*	.187**
	(.073)	(.072)	(.088)	(.068)	(.066)	(.082)	(.074)	(.074)	(.090)	(.074)	(.073)	(.089)	(.070)	(.070)	(.085)
Observations	4826	4826	4826	4839	4839	4839	4807	4807	4807	4810	4810	4810	4813	4813	4813
R-squared	.040	.040	.036	.179	.195	.167	.036	.033	.039	.041	.040	.042	.025	.026	.026

Table 16: FE estimates of the impact of the program on student outcomes, using a 6, 3 and 9 months definition of time periods.

Exam	Math			Portuguese			Biology			Physics			Rest		
Delay (months)	6	3	9	6	3	9	6	3	9	6	3	9	6	3	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
progdur	-.198***	-.206***	-.214***	-.131***	-.134***	-.169***	-.145***	-.176***	-.202***	-.139***	-.159***	-.194***	-.086*	-.111**	-.093**
	(.030)	(.030)	(.026)	(.042)	(.042)	(.036)	(.037)	(.036)	(.031)	(.036)	(.036)	(.031)	(.046)	(.046)	(.040)
progaft	-.283***	-.280***	-.352***	-.235***	-.230***	-.238***	-.231***	-.229***	-.199***	-.249***	-.250***	-.285***	-.112**	-.114**	-.123*
	(.030)	(.030)	(.046)	(.042)	(.041)	(.064)	(.036)	(.036)	(.055)	(.036)	(.035)	(.054)	(.046)	(.045)	(.070)
Observations	9661	9661	9661	9709	9709	9709	9587	9587	9587	9607	9607	9607	9541	9541	9541
R-squared	.646	.646	.646	.332	.332	.332	.415	.416	.415	.376	.373	.376	.171	.171	.171

Table 17: FE estimates of the impact of the program on student performance, using the sample without private schools.

Exam	Math			Portuguese			Biology			Physics			Rest		
Delay (months)	6	3	9	6	3	9	6	3	9	6	3	9	6	3	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
progdur	-.155***	-.161***	-.173***	-.050	-.048	-.083**	-.100***	-.130***	-.153***	-.063*	-.080**	-.112***	-.034	-.058	-.046
	(.029)	(.029)	(.026)	(.041)	(.040)	(.035)	(.035)	(.035)	(.031)	(.035)	(.035)	(.030)	(.044)	(.044)	(.039)
progaft	-.245***	-.240***	-.317***	-.142***	-.133***	-.149**	-.175***	-.172***	-.145***	-.153***	-.152***	-.194***	-.072	-.072*	-.088
	(.030)	(.029)	(.044)	(.041)	(.040)	(.061)	(.036)	(.035)	(.053)	(.035)	(.035)	(.053)	(.045)	(.044)	(.066)
Observations	7947	7947	7947	7975	7975	7975	7909	7909	7909	7904	7904	7904	7895	7895	7895
R-squared	.657	.657	.657	.387	.387	.387	.444	.444	.444	.406	.406	.406	.199	.136	.199

Table 18: FE estimates of the impact of the program on student performance, using the sample with only *on-hold* schools in the control group.

Exam	Math			Portuguese			Biology			Physics			Rest		
Delay (months)	6	3	9	6	3	9	6	3	9	6	3	9	6	3	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
progdur	-.101***	-.112***	-.112***	-.015	-.011	-.037	-.069*	-.104***	-.127***	-.003	-.024	-.053*	-.025	-.049	-.039
	(.030)	(.030)	(.027)	(.039)	(.039)	(.035)	(.036)	(.036)	(.033)	(.035)	(.035)	(.032)	(.042)	(.042)	(.038)
progaft	-.177***	-.174***	-.247***	-.084**	-.071*	-.088	-.149***	-.151***	-.122**	-.085**	-.088**	-.131**	-.067	-.067	-.085
	(.032)	(.031)	(.044)	(.041)	(.041)	(.057)	(.038)	(.038)	(.053)	(.037)	(.037)	(.051)	(.045)	(.044)	(.062)
Observations	4826	4826	4826	4839	4839	4839	4807	4807	4807	4810	4810	4810	4813	4813	4813
R-squared	.678	.678	.678	.467	.467	.467	.493	.494	.493	.477	.477	.477	.244	.244	.244