Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXV

Average and quantile effects of more instruction time in low achieving schools: evidence from Southern Italy

Direttore della Scuola: Ch.mo Prof. Monica Chiogna

Supervisore: Ch.mo Prof. Erich Battistin

Dottoranda: Elena Claudia Meroni

Acknowledgements

I would like to thanks all my XXV cycle colleagues: Manu, Mari, Clovis, Fede, Tonio e Ricardo. I'm sure that without their help and support I would have never survived the first year. Thanks also to all the other PhD students and post-doc that I met and hung out with during my year in Padova. Thanks to Anna and Angi for their friendship and company.

Thanks to my supervisor, Erich Battistin, for his precious help and for the enthusiasm and passion he has for research. He was always there when I needed his support and I've learn a lot from him. I know it is very rare to find a supervisor like him, thus I'm really grateful. Thanks also to all the other professors I met in Padova.

A special thanks goes to the Dondena people: Arnie, which introduced me to the academic research supervising my master thesis and let me work at the Dondena in the following year: probably without meeting him I would have never started the PhD. Then thanks to all the directors that always found a place for me at centre: Francesco, Vincenzo and Ross. And thanks to all the Dondena friends for their company and support during these years: Agnese, Alessia, Alice, Betta, Bruno, Elena, Emanuele, Francesco, Frank, Giulia, Sissi, Lara, Maria, Nicola and Noemi. A special thanks to Chiara, from which I really learn a lot.

Thanks to Professor Esping-Andersen for giving me the possibility to stay in his department at UPF: it was a great time of my PhD life. Thanks to my office mates in Barcelona, it was great to meet them all: Lea, Dani, Diederik, Sander and Maike. (Of course thanks also to Bruno.. for the second time..).

Thanks to all the INVALSI team: without their help I would not have had the data to complete my thesis, so they have been fundamental for my PhD. And thanks to Donatella, Pietro and Joyce, my "Roman family", that embraced me like a daughter.

Thanks to my parents, for the way you grew me. I now know that you have really done the best for me, I hope to be as good as you with the children me and Marco will have the grace to grow. And finally thanks to Marco... we got married during this PhD, between my first year exams... everybody thought I was crazy... but life has been much better with you.

Abstract

The thesis is composed by two main chapters. Both study the effectiveness of a program funded by the European Union, which was implemented during the academic year 2010/11 in low achieving lower secondary schools located in four Southern Italian regions . The intervention's aim was to increase student performances in mathematics and Italian language through the provision of extra instruction time, to be held in the afternoon, thus outside regular school time.

The first chapter focuses on average treatment effects. I control for sorting across classes using the fact that student are divided into groups distinguished by letters, they remain in the same group across grades and the composition of teachers in the school assigned to each group is substantially stable over time. I implement a difference-in-differences strategy, and compare two contiguous cohorts of sixth grade students enrolled in the same group. I contrast groups with and without additional instruction time in participating schools, to groups in non-participating schools that I selected to be similar with respect to a long list of pre-programme indicators. I find that the programme raised test scores in mathematics in schools characterised by students from less advantaged backgrounds, while no effect is found on Italian language test scores. In particular the gain is higher for the mathematical reasoning dimensions, while the knowledge of mathematics concepts is not affected.

In the second chapter, I go beyond average effects, using two non-linear methods (*Threshold difference-in-differences* and *Changes-in-changes*) which allow to recover the counterfactual distribution of the treated group had it not been treated and the quantile treatment effects of the intervention. Both methods suggest that the positive effect documented for mathematics is driven by larger effects for the best students in the group, while low achieving students seem not to benefit form the intervention.

Sommario

La tesi è composta principalmente da due capitoli. Entrambi studiano gli effetti sui risultati scolastici in Italiano e matematica di un programma finanziato dall'Unione Europea. L'intervento è stato implementato in alcune scuole medie di quattro regioni del Sud Italia durante l'anno scolastico 2010/11 e ha lo scopo di migliorare i risultati in italiano e matematica degli studenti coinvolti attraverso ore extra di lezione tenute nel pomeriggio, quindi in più rispetto al normale orario scolastico.

Il primo capitolo si focalizza sull'average treatment effect dell'intervento. Attraverso un matchig di scuole simili e una strategia di difference-in-differences, che sfrutta osservazioni ripetute di studenti appartenenti alla stessa sezione in due coorti contigue, trovo che il programma ha effetti positivi sui punteggi in matematica, solo nel gruppo di scuole caratterizzate da un profilo socio-economico basso. In particolare l'effetto è maggiore nell'ambito cognitivo, cioè l'ambito che coinvolge il ragionamento e lo sviluppo del pensiero matematico, mentre l'aspetto di pura conoscenza dei concetti matematici rimane inalterato. Sui punteggi di italiano non si trova invece nessun effetto.

Nel secondo capitolo invece identifico, attraverso due metodi diversi (il "Threshold difference-in-difference" e il "Change-in-changes"), l'intera distribuzione controfattuale del gruppo di classi trattate in assenza di trattamento, e ricavo quindi i quantile treatment effects. Con entrambi i metodi si trova che l'effetto positivo trovato nelle scuole caratterizzate da un profilo socio-economico basso, è influenzato da alti guadagni per gli studenti migliori, mentre gli studenti peggiori non sembrano beneficiare del programma.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Dating back to the mid Sixties and the publication of the Coleman Report (Coleman et al., 1966), the discussion on how to improve skills formation among students has been one of the most debated topics in social sciences. In the field of economics of education most of the attention has been devoted to measure the effect of a range of school inputs on student outcomes (typically student achievement in core subjects such as reading, mathematics and sciences), in order to better allocate resources and to reduce achievement gaps between children coming from different social background. In last two decades, a vast number of studies has tried to disentangle the effect of specific inputs in various contexts and across various grade levels; inputs such as class size (Krueger, 1999), teacher quality and training (Hanushek et al., 2005; Rivkin et al., 2005; Angrist and Lavy, 2001), instruction time (Lavy, 2012; Marcotte and Hemelt, 2008) and remedial education (Bettinger and Long, 2009; Jacob and Lefgren, 2004). In light of the growing consensus on the impact of the core elements of educational production functions, some commentators encourage to shift the efforts towards the evaluation of the effectiveness of single programs (Lavy and Schlosser, 2005; Jacob and Lefgren, 2004).

In addition, understanding the key drivers of quality in education has a fundamental role for the achievement of the Europe 2020 targets. The relevance of this problem for policy making is particularly important in areas facing marked socio-economic deprivation, and thus being at risk of lagging behind in their development. Given the conspicuous investments made by the European Union to finance structural assistance, providing evidence on the key dimensions that should be targeted by public interventions in Europe adds to the discussion on the most effective growth strategies for the coming decades. This thesis focuses on education policies that mandate low achieving students to extra hours at school, thus shedding light on the effects of instruction time on academic achievement. Despite the public opinion in the recent years has brought to the forefront the potentials of increasing instruction time, quasi-experimental evidence on this issue is still relatively scarce.

I exploit variation in the number of hours spent at lower secondary school that results from a remedial education programme implemented in selected regions of Southern Italy that are eligible to receive the EU Regional Development Funds (Objective 1 regions) and the European Social Funds. The rationale for intervening stems from the fact that schools of these areas are characterised by markedly lower student performance in the various dimensions of learning if compared to schools in the rest of the country. The Quality and Merit Project[1] (PQM in what follows) is an intervention rolled out in 2010 and targeting low achieving schools of Objective 1 regions. Participation of schools is not compulsory, and is conditional on past performance indicators. Applicant schools are ranked according to a series of indicators (i.e. percentage of retained students and drop out rates), and only those at the bottom end of the performance distribution are enrolled. Schools admitted to the programme must organise remedial education activities outside the regular school hours in a selected number of classes declared *ex ante* by school principals at the time of the application, all costs being covered by the EU structural funds.

This program is in line with other interventions implemented internationally, such as the “No Excuses” charter schools in New York, Boston and other US cities, which emphasises the importance of increasing instruction time (Dobbie and Fryer Jr, 2011).

I use longitudinal information on test scores for consecutive cohorts of students enrolled in PQM schools before and after 2010, and contrast the resulting time series to that of similar schools located in Objective 1 areas but *not* enrolled in PQM. The availability of longitudinal information for all schools in both pre and post PQM periods allows us to estimate the causal effects of extra time in education on test scores in mathematics and Italian language.

The rationale for increasing the number of hours spent by students at school lies on the direct effects of education on learning, thus the more the child is exposed to school time, the more he will learn as a comulative process; and on side-benefits coming from the lower exposure to the risk of negative behaviour (e.g. criminality, or teen-pregnancy), especially for students from low socio-economic backgrounds. However, much of the evidence on the effects of instruction time is descriptive in nature, and fails to address the possible

[1]This project is financed by the EU funding-PON Istruzione 2007-2013 (A-2-FSE-2009-2)

endogeneity arising from the spurious correlation with other school inputs and family characteristics. Different strategies have been employed over the years to deal with this problem, yielding mixed evidence.

A first strategy exploits between and within country variability in the time exposure to different subjects across EU school systems. Lee and Barro (2001) use panel data for 59 countries to assess the impact of the time spent at school during the year on student performance, finding no effects on test scores. Using TIMMS data for 39 countries, Wößmann (2003) finds that the effect of instruction time is positive and significant, but negligible in size. Using a sample of students from more than 50 countries that participate in PISA, Lavy (2010) finds that instruction time has a positive and significant effect, though negligible in size, on test scores. Mandel and Süssmuth (2011) exploit cross state variation in instruction time within Germany and find that instruction time by subject, measured in cumulative terms, is a highly robust determinant of cognitive achievement.

A second strategy exploits the exogenous variation on length of school year that results from different quasi-experimental settings. Marcotte (2007) and Marcotte and Hemelt (2008) consider the variation in school-closing days for snowfalls in Maryland, finding that students perform better in years with less unscheduled closing days. Hansen (2008) also exploits weather-related cancellations in Colorado and Maryland, as well as change in test-date administration in Minnesota, which moved 5 times in 5 years. The results point to positive effects of the number of school days on student performance. Sims (2008) uses a similar idea exploiting a reform in Wisconsin, finding that additional school time is associated with a small increase in mathematics for fourth grade students, but does not affect reading competencies. Pischke (2007) exploits the variation in instruction time resulting from the German "short school years", finding that shorter years are associated with an increase in grade repetition. Bellei (2009) finds that the Chilean full school day programme has been beneficial for both reading and mathematics test scores. Full school day compared to half school day was found to have a positive effect on learning outcomes also in Kindergarten (Robin et al., 2006; DeCicca, 2007; Lash et al., 2008; Gibbs, 2010).

A different stream of literature, which is closer in spirit to the intervention that I consider in this thesis, investigates the effect of a longer school time but conceived as more hours per day at school rather than more school days per year. Extra-education is organized by opening schools for longer hours during the afternoon, either providing extra-instruction time on curricular activities or helping students from less advantage backgrounds doing their homework. Lavy and Schlosser (2005) reports quasi-experimental estimates of the effect of a programme providing targeted additional instruction time to low achieving high school students in Israel. The analysis documents an increase in college matriculation rates of about 3 percentage points. Zimmer et al. (2010) considers Pittsburgh Public Schools, which enacted various initiative to improve student performance via extra education and tutoring initiatives. Using longitudinal data on students, the authors document positive effects for mathematics but not for reading. Lavy (2012) exploits a school finance policy experiment undertaken in elementary schools in Israel that changes the length of the school week and the time allocation to core subjects. He finds that spending more time at schools and on key tasks yields an increase of achievement in mathematics, English and sciences; and the effect is much larger for students coming from low socio-economic background and in school whose students have homogenous socio-economic background. Other programmes, however, were found to be ineffective: this is the case of the programmes evaluated by Dynarski et al. (2004), Checkoway et al. (2011) and Meyer and Van Klaveren (2011).

The literature on the effects of increased instruction time on academic performance overlaps, to a large extent, with that considering the effects of specific remedial programmes targeting low achieving students. In many instances, students benefitting from increased time at school are those from less advantaged backgrounds, for whom extra activities at school often come in the form of remedial classes. The intervention considered in this thesis shares with remedial education programmes the idea that public investment should target the lower end of the performance distribution. However, rather than targeting only low achieving students in the class, the PQM programme targets all students in low performing schools in the most deprived areas of the country. Thus we can think of PQM as an intervention providing remedial education to the most needing schools.

In the stream of the literature studying the effects of remedial education, Aiken et al. (1998) find a positive effect of a university-level freshman remedial writing programme and Bettinger and Long (2009) identify positive effects of remediation on college outcomes of under-prepared college students in Ohio. Positive results of remedial high school programmes were found also for courses targeting younger students (Lang et al., 2009) and implemented outside the US context (Lavy and Schlosser, 2005). On the other hand, Calcagno and Long (2008) find that remedial courses increase the probability of completing the first year of college of a sample of more than $100,000$ students in Florida, but they do not affect the likelihood of completing the whole degree. A previous randomised study on a summer school programme including summer employment and academic remediation aimed at contrasting early school dropouts, found no effects (Grossman et al., 1992). The effect of remedial classes in Italy has been recently studied on a sample of students in upper secondary schools (Battistin and Schizzerotto, 2013), finding a positive effect of remedial classes on the performances of the academic track students, while it has a negative effect on students attending vocational high schools.

1.2 Main contribution

The main contributions of this thesis can be summarised as follows. First, I create a new variable measuring students performances which resemble test scores used in international survey; indeed the variable provided by the INVALSI (National Institute for the Evaluation of the Educational System) throughout the national assessment test is just the percentage of correct answers in mathematics and Italian language. I develop a weighting scheme that allows to give more weights to harder questions, and less weight to easier questions, so to create a measure of the outcome which is, a least on a theoretical point of view, comparable with international measures of student performances.

Second, I implement a novel strategy to control for sorting of students across classes that builds upon features of the Italian school system. Students enrolled in lower secondary schools are divided into groups, called *sezione*, distinguished by letters, and remain in the same group (*sezione*) for the whole cycle of studies.[2] Assignment of students to groups is not random, and results from idiosyncratic criteria followed by school principals and from the parents' pressure to have their children in the most prestigious *sezione* of the school. The key feature that we exploit for identification is that prestige depends on the quality of teachers, and that it is common practice in Italian schools to observe the same teachers in the same *sezione* over the years and across grades. I make the assumption that sorting of students across groups is stable over time, and compare changes in test scores for consecutive cohorts enrolled in the same *sezione* before and after PQM.

Third, I exploit within school variability in the enrolment of classes to assess the existence of indirect effects on test scores of students not directly involved in the PQM programme. The identifying source of information exploited comes

[2]Normal age for enrolment in lower secondary education is 10, and progression to the upper secondary level - which is compulsory by the Italian law - is expected three years later. To give an example, consider a school composed of 6 classes: 2 6th grade classes, 2 7th grade classes and 2 8th grade classes. This school will have 2 *sezioni*, which we call A and B. Hence, in each year there will be *sezione* A class and *sezione* B class of 6th graders; *sezione* A class and *sezione* B class of 7th graders; and *sezione* A class and *sezione* B class of 8th graders. A student assigned to *sezione* A class in 6th grade in year 1, will be, with the same peers, in *sezione* A class in 7th grade in year 2, and so on.

from the comparison between students in control classes of PQM schools, and students in all other schools located in Objective 1 areas.

Fourth, I go beyond averages and investigate the extent of heterogeneity in the effects of instruction time along two different dimensions. On the one hand, we allow for returns that depend on the number of school hours that come on top of normal school time. Given the assumptions that we discuss further below, we can benefit from a good deal of variability in this measurement across PQM classes. On the other hand, we combine difference in differences estimation with quantile regression analysis using the methodology in Firpo et al. (2009) and further extended to the difference-in-differences setting by Havnes and Mogstad (2010) and the methodology in Athey and Imbens (2006), to identify quantile treatment effects of the intervention.

My main results can be summarised as follows. First, I find that the PQM programme has had a positive effect on average test scores in mathematics but *not* in language. This effect is driven by large average returns to participation only for students in the most problematic schools, that is schools in the lowest tertile of student achievement in the pre-programme period. These are learning environments characterised by the highest retention rates and whose students come from markedly less advantaged backgrounds. In addition the positive effect is significant only in the part of the test measuring mathematical reasoning, and not mathematical knowledge, suggesting that the extra instruction time in the afternoon does not add much in terms of knowledge of mathematical concepts, but can help students boosting their abilities to think and and to apply their knowledge.

Second, for schools in the top tertile we find that extra hours tailored around reading activities have had a *negative* average effect on test scores in mathematics, and no effect on language. Given that language abilities are found to be less responsive to PQM across learning environments, we interpret this result concluding that in the least problematic environments instruction time should target activities that enhance mathematical abilities, as the additional time spent at school engaged in reading activities may substitute the time that students would have invested on mathematics.

Third, because of the importance of distributional effects, we go beyond averages and assess how PQM has affected achievement across *quantiles* of the test score distributions. I maintain the stratification by school tertile to understand the interplay of instruction time with the learning environment. I find that the average returns to PQM documented for the lowest tertile conceal sizeable effects after the 40th percentile of the test score distribution in mathematics. The absence of average causal effects for students in all remaining schools translates into the same conclusion for the various quantiles considered. Moreover, I find that the negative effect of reading activities on mathematics for schools in the top tertile is concentrated at the top end of the test score distribution. In other words, extra hours spent at school by students on reading come at the cost of outstanding performance in mathematics.

The thesis is structured as follow: in Chapter 2 I provide background information on the Italian school system and detailed description of the PQM intervention and of the data used; in Chapter 3 I provide evidence on the effectiveness of the program, focusing on average treatment effect; in Chapter 4 I go beyond average effects and estimate the effect of PQM across *quantiles* of the test score distributions. Finally I provide some conclusion and suggest possible policy implications emerging from the evaluation of the program.

Chapter 2

Description of the intervention in the Italian context

2.1 The Italian school system

In the italian school system students attend primary school from grade first to fifth, then lower secondary school, from grade sixth to eight. The school programmmes taught in primary and lower secondary schools are settled by the Italian Ministry of Education, hence being identical across the whole country. At the end of the eight grade students start higher secondary school and are free to choose among three main different major tracks: vocational high school (Istituto professionale and Corsi di formazione professionale), technical high school (Istituto tecnico) and academic high school (Liceo).

At the beginning of each block (primary, lower seconday and higher secondary) students are assigned to a specific class, which is called *sezione*, and they remain in the same class for all the length of the block (i.e. 5 years in the primary school, 3 years in the lower secondary school, and 5 years in the secondary school). This implies that once a student is assigned to a class (*sezione*) he will follow all the subjects with the same peers for all the years of the block. To provide an example, assume that a given lower secondary school is composed by a total of 6 classes: 2 sixth grade classes, 2 seventh grade classes and 2 eight grade classes. This school has 2 *sezioni*, which we call A and B. Hence each year there will be a class of sixth graders *sezione* A, a class of sixth graders *sezione* B; a class of seventh graders *sezione* A, a class of seventh graders *sezione* B; a class of eight graders *sezione* A, and a class of eight graders *sezione* B. A student which is assigned to *sezione* A in sixth grade in academic year 1, will be, with the same peers, in*sezione* A, in seventh grade in academic year 2, and so on.

In theory assignment, of both teachers and students, to the different *sezioni* shuold be random, but in practice it is well known that there are some mechanisms (parents' pressure to have their children in a given *sezione*, school principal assigning some teachers to a given *sezione*, ...), which could lead to a different composition of the different *sezioni* inside a school. Nevetheless it is quite common that a teacher is assigned to the same *sezione* throughout the years and across grades and, if there exist an assignment mechanism to a *sezion* based on students' ability, this mechanism is constant through time. Therefore if all the best students, or the best teachers are assigned to *sezione* A in year t, it is very likely that the same mechanism would be implemented in year $t+1$. In Table 2.1 I report the Kendall Tau rank correlation coefficient for the performances in mathematics and language of two consecutive cohorts of students belonging to the same *sezione* across two consecutive years. Both the coefficients shows a positive and significant relationship, meaning that rank similarity of the same *sezione* across years is a plausible assumption.

Table 2.1: Kendall's tau-a correlation coefficients for the rank of *sezione* across years

Kendall's tau for the rank in Italian language	0.287^{***}
Kendall's tau for the rank in mathematics	0.293^{***}
Number of *sezione*	595

The correlation coefficient is estimated using the rank of a *sezione* in year 2009/10 and year 2010/11. Only schools chosen as control have been considered (See Chapter3)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.2 The regional differences between the North and the South

Systematic evidence from international surveys (IEA-PIRLS 2006; IEA TIMMS 2007; PISA 2003, 2006 and 2009) has identified the gap between the Italian school system and that of other OECD countries. It is now well documented that Italian students perform below the European average in both mathematics and reading. This figure conceals a good deal of variability across regions, with Northern areas performing in line with other European countries and Southern areas performing markedly below. The recent experience on national assessment tests has demonstrated that, while the North/South divide is contained for second graders, it increases at the end of the primary school and grows even larger in middle schools (INVALSI, 2010b). For these reasons, four regions located in the Objective 1 area (Campania, Sicily, Calabria and Apulia) are eligible to benefit from EU Regional Development Funds and from the European Social Fund, for the period 2007/13, to improve teaching and learning processes in middle and high schools. One of the actions taken with these fundings was the implementation of the PQM programme.

There is evidence that regions that have eventually employed EU funds have sensibly improved their performance at PISA tests between 2006 and 2009, in particular in Apulia and Sicily (INVALSI, 2010a).[1]. For example, according to the distribution of test scores in mathematics for 2006 Apulia was ranked seventeenth amongst the 20 Italian regions, and eleventh in 2009. The same figures for reading are sixteenth in 2006 and twelfth in 2009, and for sciences are seventeenth in 2006 and thirteenth in 2009. Taken at face value, this result may be suggestive of possible causal effects at work, and is currently animating the public debate on the effectiveness of EU investments in the most deprived areas of the country. The lack of empirical evidence on this issue, for Italy and more in general for the optimal design of public policies aiming at EU 2020 objectives, is a gap that this thesis intends to fill. The only available evidence of the effect of the EU funds in Italy is given by Falzetti et al. (2012): comparing schools participating in the PISA test both in 2006 and in 2009, they find that schools belonging to the regions eligible for receiving the EU funds, compared to schools belonging to regions not eligible, but located in the South of the country, improved their performances from year 2006 to 2009. Given the lack of availability of standardised test scores taken by all the schools in the country before 2009/10, my thesis is one the first begin able to provide a rigours evaluation of the effect of the EU funds on a wider scale.

2.3 The PQM Program

The PQM programme targets lower secondary schools in the four Obejctive 1 regions. It was first implemented in the academic year 2009/10, subsidising additional hours in mathematics in 215 schools. In the following academic year, new schools were added along with the possibility of extending instruction time to Italian language. The total number of schools involved in the academic year 2010/11 was 223, of which 84 already participated in the previous year. In either rounds, participation was not compulsory: applicant schools were enrolled giving preference to those performing at the lower end of the distribution according to the percentage of retained students and drop out rates. The criteria used for admission were the same in both years.

[1]Calabria was not sampled in the PISA 2006, thus this does not hold for this region

Schools enrolled organise extra activities outside regular hours, in a selected number of classes (two per subject). At the time of the application the school principal has to point out the two teachers[2] that will provide the extra education, ad thus the corresponding two classes that will be treated. Teachers are pointed out since a part of the intervention foresees that teachers of the selected classes undertake a training course, whose aim is to help them organise the extra activities that they will hold in the afternoon. The training consists of 60 hours (30 hours of formal training and 30 hours online) and it helps the teachers to set up a *Improvement Plan*, based on the return of the results of the standardised test which treated classes take at the beginning of the academic year (October). This test should help teachers targeting pupils who are in need and areas on which intervene. The training is held in groups of 10 teachers (i.e. 5 schools), and it is supervised by a mentor who provide support in respect of their decisions about how to organise remedial and extra activities during the school year. It is important to stress that the training is not content focused, thus it does not affect the teachers' competences and their knowledge in the subjects, but it simply supports them in the decision on how to organise the extra activities and it provides them with some material that can be used during such activities.

The afternoon activities planned per class can range from 1 to 8, and teachers receive extra-salary for their extra loads.[3] Each activity foresees an average of 15 hours of extra education to be held outside the regular school time to students, and the teacher is free to decide how many activities and how many students to involve. In our data, the average number of students involved as a proportion of class size varies between 25 percent and 100 percent, nevertheless in more than 75 percent of classes at least 50 percent of students participate in the afternoon activities. In most classes (about 65 percent) the number of activities chosen is between 2 and 4. This corresponds, on average, to additional 30 to 60 hours spent at school by participating students over the school year.

2.4 Data

Data at the school level are provided by the Italian Ministry of Education, through INVALSI. This administrative data provides general information about the schools characteristics (number of students, student to teacher ratio, drop out rates, ...) and the exact municipality where the school is located, thus

[2]The only requirement set for teachers is that they should be permanent teachers.

[3]Teachers receive 50 euro per hours gross, thus considering their salary, planning 4 activities would make one month salary.

geographical and demographic characteristics of the environment where the schools operate are also available.

Data at the student level are collected directly by the INVALSI, which is in charge of testing the Italian students' performances trough a national assessment test in mathematics and language. This test was introduced on a small sample of schools in second and fifth grade in academic year 2007/08, and since the academic year 2009/10 it is taken by all students in the country at the end of second, fifth, sixth, and eight grade. The data contain information on the results of the standardised tests, both for mathematics and language, the main socio-demographic characteristics of the child and his family (gender, year of birth, origin, level of education and employment status of the parents, household composition) and questions about motivation and perception of the school.

In addition I have a unique piece of information about the class (*sezione*) in which each student is enrolled; information that as previously explain we use to control for sorting of students and teachers into the different classes.

The two datasets, the one about the schools and the one about the students, can be merged trough a unique code which identify schools. So, for each students we not only know his personal characteristics and test scores, but also all the general characteristics of the school where he is enrolled and the exact municipality where he is living. On the other side, for each school we are able to reconstruct average performances and average characteristics of the students enrolled in sixth and eight grade.

Language test are built to measure reading proficiency (in particular the ability of the students to understand and interpret a text) and lexical and grammatical knowledge, while mathematics test is measuring knowledge of the mathematics contents and logical and cognitive processes used in the mathematical reasoning. The tests are composed mainly by multiple choices questions, in which the students have to select the right answer out of two or four possibilities; in mathematics there are also few open questions. The score provided by the INVALSI is calculated simply as percentage of corrected answers out the total number of questions (42, in 2010 and 43, in 2011 for mathematics and 58, in 2010 and 82, in 2011 for language), and hence varies between 0 and 1.

While this variable gives an insight about students' achievement, it lacks of a fundamental dimension since it doesn't not take into account the different level of difficulty of each question, and it gives equal weight to very hard questions and to very easy questions. Thus a student answering correctly to the 20 easiest questions will have the same score of a student answering correctly to the 20 harder questions. In order to overcome this issue and to construct an outcome variable which could resemble test score used in international tests, such as the PISA[4] I construct a weighting scheme that gives more weight to harder questions. For each question I calculate the proportion of students among all the students in the four Objective 1 regions (Apulia, Calabria, Campania and Sicily), who answered correctly, therefore the large is the proportion the easiest is the question. Weights are then constructed simply as the inverse of this proportion, with higher weights assigned to harder questions. The outcome variables we construct is the weighted average of the number of correct answers in mathematics and language, and they have been standardised so to have in each each year mean 0 and standard deviation 1.

In the analysis I use as outcome variable both the original variable provided by the INVALSI, and I refer to it as percentage of correct answers, and the variable I constructed taking into account different level of difficulty in each question, and I refer to it as test score.

Knowledge of mathematics is assessed by considering two dimensions: a *content dimension*, specifying the subject matter (numbers, space and shapes, data and forecast, and functions), and a *cognitive dimension*, measuring the mental process employed when engaged with the content. Each question in the test is explicitly designed by the INVALSI to measure two mutually exclusive cognitive domains: *knowledge* (which refers to the student's knowledge of facts, concepts, tools, and procedures in mathematics), and *reasoning* (which focuses on the student's ability to apply knowledge and conceptual understanding in a problem situation).[5] Similarly, the test for Italian language is designed to measure reading proficiency (i.e the ability to understand and interpret a text) and grammatical knowledge. Since each question in the test can be mapped into one of the above domains, in my analysis I will distinguish between outcomes that refer to Italian language (comprising *reading comprehension* and *grammatical knowledge*) and mathematics (comprising *mathematical knowledge* and *mathematical reasoning*). Standardised test scores will be considered

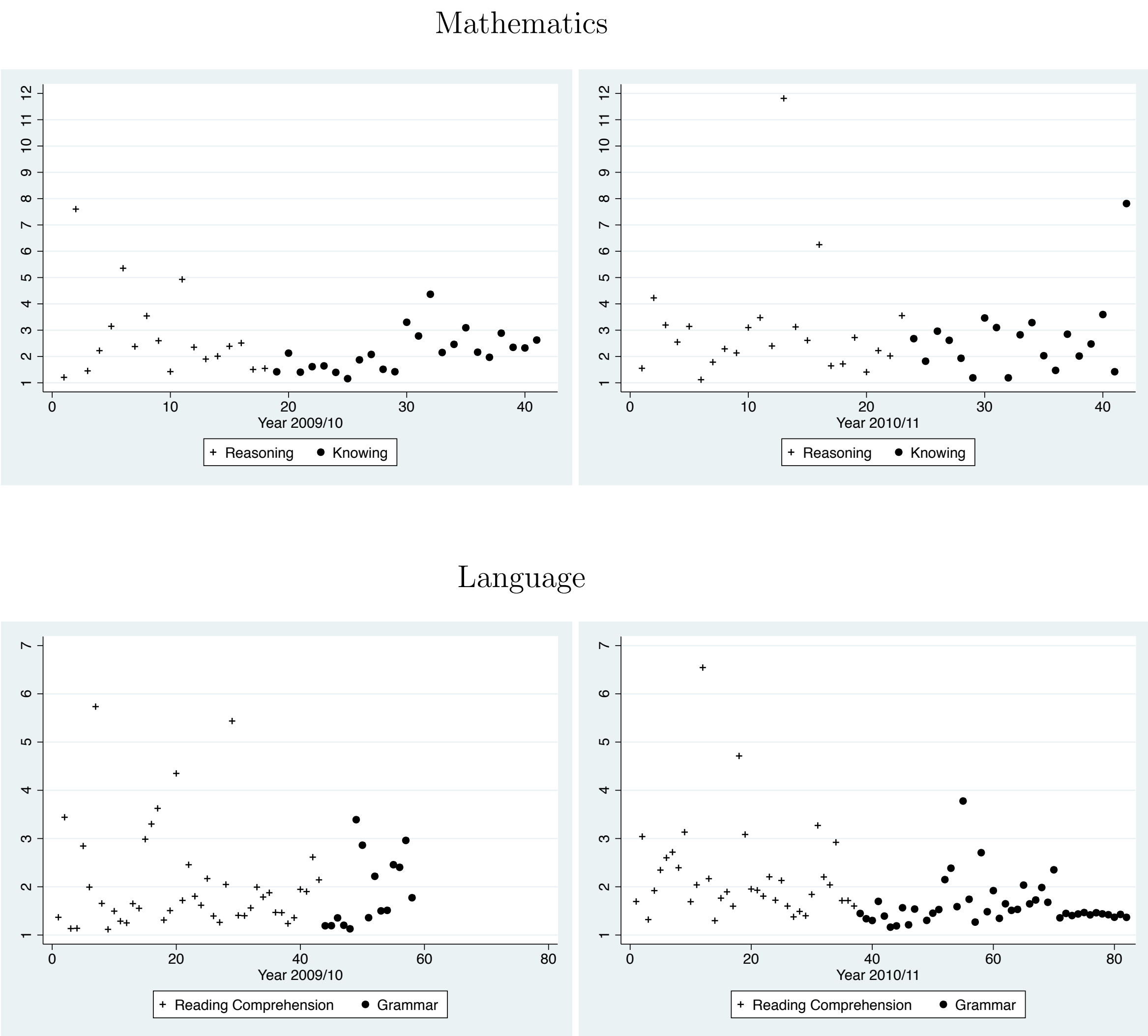
[4]PISA test questions are divided into levels, with Level 1 questions requiring only most basic skills and increasing difficulty in each level, thus the PISA score takes into account of the different level of difficulty of each question.

[5]The INVALSI proposes 8 different subcategories, which can be obtained from questionnaires. The knowledge domain is composed of three categories: 1) Knowing the specific mathematical contents, such as mathematical objects, properties, structures; 2) Knowing algorithms and procedure in geometry and arithmetic; 3) Knowing different ways of mathematics expression: verbal, written, symbols, graphical. The reasoning domain is composed of 5 categories: 1) Being able to solve problems using mathematical tools; 2) Being able to identify objects' measurability and being able to use measurements tools; 3) Acquire and use mathematical thinking; 4) Using mathematics to deal with information coming from the science, technology, economic and social fields; 5) Being able to recognise shapes in the space.

throughout.

The distribution of weights in the different domains is reported in Figure 2.1, mathematics is in the top panel and language in the bottom panel. The average weight for *mathematical reasoning* is 2.78 in 2009/10 (18 questions) and 3.05 in 2010/11 (23 questions), while for *mathematical knowledge* is 2.18 in 2009/10 (23 questions) and 2.68 in 2010/11 (20 questions). Weights appear to be slightly higher for the former domain. The average weight for *grammatical knowledge* is 1.90 in 2009/10 (15 questions) and 1.63 in 2010/11 (45 questions), while for *reading comprehension* is 2.07 in 2009/10 (43 questions) and 2.24 in 2010/11 (37 questions). We still observe enough variability in the level of difficulty across questions.

Figure 2.1: Weight assigned to each question in Mathematics (*reasoning* and *knowing*) and Italian Language (*reading* and *grammar*) domains



Chapter 3

Average effects of extra instruction time on student achievement

3.1 Selection of the relevant sample and descriptive statistics

I will focus only on the *second* wave of PQM, which was implemented in the year 2010/11. These choice is very pragmatic, and driven by the data problems related to the participation of schools in the national assessment test in 2008/09. Thus I decided to consider 2009/10 as the pre-programme period, and employ a difference-in-differences strategy that makes use of test scores for the following waves. I also decided to drop from the analysis schools who were participating in the programme in both years, thus concentrating just on schools who were selected in 2010/11 for the first time. In addition, I drop from the sample the schools who were enrolled in the programme in the first year, but not in the second, since it would not be correct to include them in the control sample of schools.

In Tables 3.1, 3.2, 3.3 and 3.4 we find descriptive statistics for all the schools in the four regions. The number of PQM schools varies between the four regions[1], with just 11 schools in Calabria and around 40 in the other three regions. The tables reveal a differential process of selection of the school inside each region. In Campania (table 3.1), the two group of schools are not dissimilar on wide range of variables, except for the student to teacher

[1]These numbers take into account that we dropped all the schools who were doing the PQM program also in the pre-treatment year, 2009/10

ratio and the school size: PQM schools are bigger and with a higher student to teacher ratio. In Apulia PQM schools are bigger and perform worst than non-PQM schools, indeed percentage of correct answers in PQM schools is on average 2 percentage points less in both mathematics and Italian language; PQM schools, moreover, show a higher rate disable students than non PQM schools. In Calabria PQM schools have an higher proportion of permeant teachers and of student to teacher ratio, and PQM schools are bigger than non PQM and are located in larger towns; in addition a lower proportion of students is attending more than 30 hours per week. Finally, in Sicily the only two significant difference regards fundings for students' activities and location of the schools. What seem to be common among the four region is school size, PQM schools are bigger than non PQM schools, probably because of the requirements (they had to be at least two permanent teachers) and the fact that PQM schools seem to be located in larger towns. Surprisingly none of the criteria used to chose PQM schools (retention and drop out rates) is different between selected and non selected schools. This could mean that groups of applicant schools on average was not characterised by high retention and drop out rates, and thus that there exist among the non applicant, and therefore not selected, some schools that have similar observable characteristics of the enrolled ones.

In order to control for the bias resulting from the non random selection of the school, I choose a group of control schools among all the non PQM schools, which share similar observable characteristic with the schools enrolled in the programme. Through propensity score matching we find a *matched pair* comparison of similar schools located in the the same province, but with different status with respect to enrolment in PQM. The propensity score is calculated separately in each region, and the matching is done one-to-one with replacement. The matching procedure along the dimension considered did not yield to any common support problem.

Variables use for the calculation of the propensity score are: average percentage of correct answers in mathematics and language in sixth grade; student to teacher ratio, proportion of permanent teachers, drop out rate, failing rate, proportion of repeating students, proportion of immigrant students, proportion of disable students, proportion of female students, proportion of students attending more than 30 hours per week, number of students, whether the school has received in the previous year other PON funds for other activities, population in town and whether school is located on a mountain municipality. Since all the school chosen to participate to PQM were public schools, before calculating the propensity score I dropped non public schools from the sample.

In table 3.5 I report the estimates of the four logistic regressions made in

Table 3.1: Descriptive statistics for PQM and non PQM schools, Campania

	(1) PQM	(2) Non PQM	(3) Difference
Mathematics, percentage of correct answers	0.578	0.582	-0.004
			(0.010)
Italian, percentage of correct answers	0.489	0.486	0.003
			(0.014)
Proportion of permanent teachers	0.917	0.890	0.027
			(0.016)
Student-teacher ratio	9.933	9.038	0.895
			(0.380)
Number of students in the school	430.3	334.7	95.54
			(35.89)
Proportion of immigrant students	0.022	0.025	-0.002
			(0.004)
Proportion of disable students	0.031	0.034	-0.003
			(0.004)
School drop out rate	0.001	0.002	-0.000
			(0.001)
School rate of failing students	0.041	0.040	0.001
			(0.007)
School rate of repeating students	0.038	0.037	0.002
			(0.007)
Proportion of female in the school	0.492	0.478	0.0148
			(0.012)
Proportion of students doing more than 30 hours	0.394	0.384	0.009
			(0.068)
School received PON funds	0.930	0.859	0.072
			(0.055)
Municipality located on mountain	0.233	0.309	-0.076
			(0.073)
(Log) Population in town	10.12	9.961	0.159
			(0.298)
Number of PQM schools	43		
Number of non PQM schools	460		

Presented in the table are descriptive statistics for the whole sample of schools in Campania. Column (1) refers to schools participating in the programme; column (2) refers to schools non participating; column (3) reports the difference between column (1) and column (2), together with the corresponding standard error (in parenthesis).

Table 3.2: Descriptive statistics for PQM and non PQM schools, Apulia

	(1) PQM	(2) Non PQM	(3) Difference
Mathematics, percentage of correct answers	0.578	0.597	-0.019
			(0.010)
Italian, percentage of correct answers	0.478	0.503	-0.025
			(0.011)
Proportion of permanent teachers	0.887	0.901	-0.014
			(0.017)
Student-teacher ratio	10.343	10.311	0.032
			(0.339)
Number of students in the school	438.9	349.5	89.43
			(32.10)
Proportion of immigrant students	0.028	0.027	0.001
			(0.005)
Proportion of disable students	0.032	0.026	0.005
			(0.003)
School drop out rate	0.003	0.002	0.001
			(0.001)
School rate of failing students	0.038	0.033	0.006
			(0.006)
School rate of repeating students	0.040	0.032	0.008
			(0.005)
Proportion of female in the school	0.480	0.484	-0.004
			(0.010)
Proportion of students doing more than 30 hours	0.262	0.314	-0.052
			(0.063)
School received PON funds	1.000	0.960	0.040
			(0.030)
Municipality located on mountain	0.326	0.186	0.140
			(0.066)
(Log) Population in town	10.61	9.969	0.644
			(0.221)
Number of PQM schools	43		
Number of non PQM schools	253		

Presented in the table are descriptive statistics for the whole sample of schools in Apulia. Column (1) refers to schools participating in the programme; column (2) refers to schools non participating; column (3) reports the difference between column (1) and column (2), together with the corresponding standard error (in parenthesis).

Table 3.3: Descriptive statistics for PQM and non PQM schools, Calabria

	(1) PQM	(2) Non PQM	(3) Difference
Mathematics, percentage of correct answers	0.596	0.560	0.037
			(0.020)
Italian, percentage of correct answers	0.503	0.479	0.024
			(0.029)
Proportion of permanent teachers	0.924	0.814	0.110
			(0.047)
Student-teacher ratio	9.409	7.486	1.923
			(0.734)
Number of students in the school	380.4	221.7	158.7
			(51.35)
Proportion of immigrant students	0.035	0.037	-0.002
			(0.012)
Proportion of disable students	0.025	0.030	-0.004
			(0.006)
School drop out rate	0.005	0.004	0.002
			(0.004)
School rate of failing students	0.037	0.045	-0.007
			(0.015)
School rate of repeating students	0.051	0.038	0.013
			(0.013)
Proportion of female in the school	0.477	0.476	0.001
			(0.027)
Proportion of students doing more than 30 hours	0.242	0.527	-0.285
			(0.142)
School received PON funds	0.900	0.877	0.023
			(0.106)
Municipality located on mountain	0.400	0.626	-0.226
			(0.157)
(Log) Population in town	10.47	9.016	1.457
			(0.431)
Number of PQM schools	10		
Number of non PQM schools	227		

Presented in the table are descriptive statistics for the whole sample of schools in Calabria. Column (1) refers to schools participating in the programme; column (2) refers to schools non participating; column (3) reports the difference between column (1) and column (2), together with the corresponding standard error (in parenthesis).

Table 3.4: Descriptive statistics for PQM and non PQM schools, Sicily

	(1) PQM	(2) Non PQM	(3) Difference
Mathematics, percentage of correct answers	0.546	0.554	-0.008
			(0.011)
Italian, percentage of correct answers	0.457	0.467	-0.010
			(0.015)
Proportion of permanent teachers	0.855	0.843	0.012
			(0.021)
Student-teacher ratio	8.655	8.594	0.061
			(0.376)
Number of students in the school	356.3	299.2	57.10
			(30.05)
Proportion of immigrant students	0.029	0.034	-0.005
			(0.009)
Proportion of disable students	0.044	0.040	0.004
			(0.005)
School drop out rate	0.005	0.004	0.001
			(0.002)
School rate of failing students	0.069	0.065	0.004
			(0.009)
School rate of repeating students	0.064	0.071	-0.007
			(0.010)
Proportion of female in the school	0.504	0.484	0.020
			(0.012)
Proportion of students doing more than 30 hours	0.371	0.407	-0.035
			(0.070)
School received PON funds	0.976	0.874	0.103
			(0.052)
Municipality located on mountain	0.286	0.449	-0.163
			(0.080)
(Log) Population in town	10.51	10.16	0.356
			(0.280)
Number of PQM schools	42		
Number of non PQM schools	419		

Presented in the table are descriptive statistics for the whole sample of schools in Sicily. Column (1) refers to schools participating in the programme; column (2) refers to schools non participating; column (3) reports the difference between column (1) and column (2), together with the corresponding standard error (in parenthesis).

Table 3.5: Probability of being a PQM School

	Campania	Puglia	Calabria	Sicilia
Italian, percentage of correct answers	-0.083	0.044	0.096	-0.007
	(0.050)	(0.066)	(0.123)	(0.045)
Mathematics, percentage of correct answers	0.031	-0.080	-0.029	-0.007
	(0.034)	(0.060)	(0.074)	(0.034)
Proportion of permanent teachers	3.221	-1.676	5.647	-0.216
	(2.330)	(2.212)	(5.161)	(1.619)
Student-teacher ratio	0.087	0.019	-0.278	-0.194
	(0.122)	(0.152)	(0.275)	(0.141)
Number of students in the school	0.002	0.002	0.003	0.002
	(0.001)	(0.001)	(0.002)	(0.001)
Proportion of immigrant students	1.997	4.740	0.330	-1.251
	(7.159)	(6.334)	(13.99)	(4.661)
Proportion of disable students	-4.468	18.30	-19.43	1.715
	(12.05)	(13.29)	(30.53)	(8.124)
School drop out rate	-3.339	9.172	24.80	9.974
	(22.34)	(34.84)	(27.56)	(12.59)
School rate of failing students	-0.569	-5.865	-20.27	3.937
	(5.453)	(8.132)	(14.28)	(3.771)
School rate of repeating students	1.275	3.061	21.47	-7.006
	(5.698)	(8.235)	(12.49)	(4.279)
School received PON funds	0.828		-0.548	1.820
	(0.632)		(1.286)	(1.044)
Proportion of female in the school	3.348	-2.505	-1.491	39.34
	(2.620)	(3.376)	(6.531)	(32.71)
Municipality located on mountain	-0.176	0.774	-1.008	-0.837*
	(0.449)	(0.422)	(0.775)	(0.401)
Proportion of students doing more than 30 hours	0.594	-0.143	-1.449	-0.163
	(0.454)	(0.638)	(1.136)	(0.500)
(Log) Population in town	-0.117	0.231	0.444	0.135
	(0.138)	(0.165)	(0.353)	(0.142)
Constant	-4.839	-1.944	-12.85*	-13.79
	(3.423)	(3.399)	(6.343)	(8.700)
Number of schools	503	286	237	450

Presented in the Table are the estimates for the four logistic regressions used to calculated the propensity score. Estimates are at the school level, using pre-program characteristics and the four columns correspond to four different regressions. In order to reach better balance, in Sicily also the variable "proportion of female squared" was included. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the four regions to calculate the propensity scores: in each region, I estimate the probability of being selected for the PQM programme, conditional on the observable characteristics at the school level. As previously seen, since PQM schools were not so different from non PQM schools, none of the variable is ever significant in the logistic equation[2].

Once obtained the propensity score, we matched each PQM school with the non treated school located inside the same province with the closer propensity score. In table 3.6 we find the final numbers of schools, classes and students in PQM and schools chosen as control, in both pre and post treatment year. The final sample of treated schools is composed by 23 schools enrolled only in PQM mathematics, 37 schools enrolled only in PQM Italian language, and 74 schools enrolled in both components of PQM. This corresponds to 122 classes receiving extra education in mathematics, 141 in Italian language and 39 in both subjects during the academic year 2010/11. Thus, although the number of schools selected for both programme was high (72 schools out of 134), then just few classes were selected to participate in both part of the programme, and this happened in smaller schools, therefore was probably due to practical reasons (not enough classes to implement the programme in 4 different classes). In Figure 3.1 I present a map of the 4 regions involved and the location and number of PQM and control schools in each municipality.

Table 3.6: Number of schools, classes and students

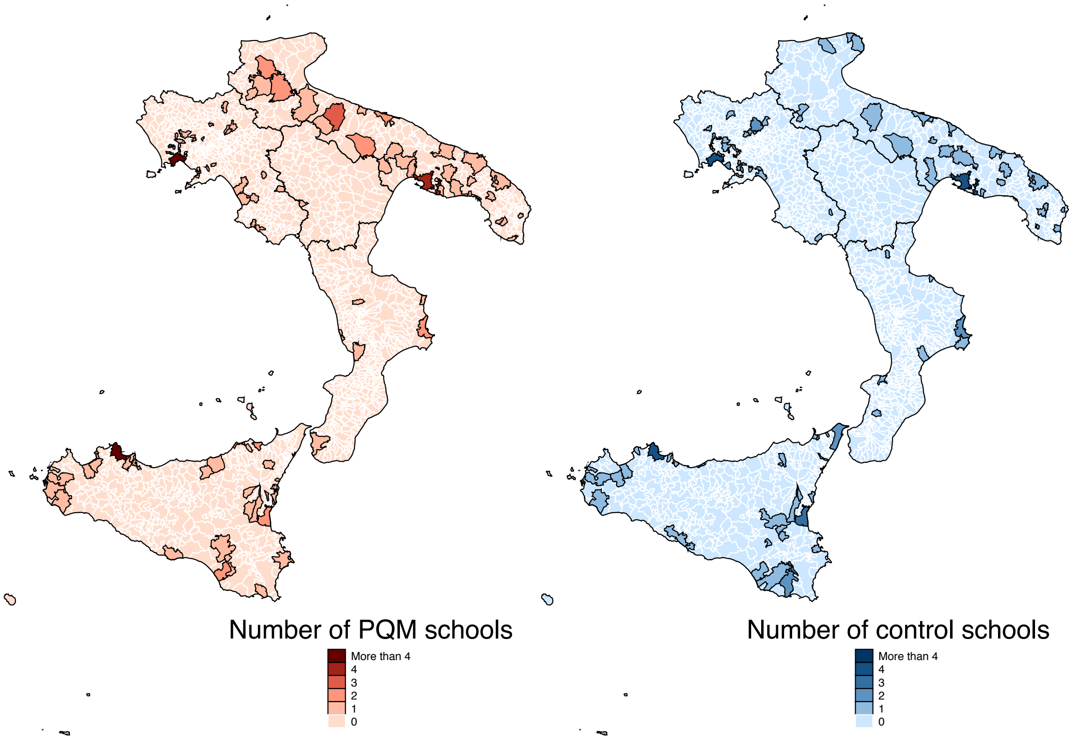
	Pre treatment year	Post treatment year
PQM schools	134	134
Treated classes	302	302
Treated students	6215	5998
Control classes in PQM schools	414	414
Students in control classes in PQM schools	8542	8412
Control schools	114	114
Control classes in control schools	595	595
Students in control schools	12455	12672

Presented in the Table are the numbers of students, classes and schools in my working sample.

Using school identifiers provided by the INVALSI, we were able to link data for the same school in the 2009/10 (pre-programme) and the 2010/11 (post-programme) year. Moreover, as discussed in the Introduction, I obtained identifiers for the group to which students are assigned at school (*sezione*). Thus my working sample consists of two consecutive cohorts of 6th graders enrolled in the *same* school and in the *same* group one one year before and one year after the introduction of PQM. Selectivity issues arising from the self-selection of

[2]Exception made for the variable being located on municipality on mountain.

Figure 3.1: Number and location of PQM and control schools in the different municipalities of the four Objective 1 regions



schools into PQM are addressed by means of school fixed effects, which are imposed after having matched participating and non-participating schools along a large set of pre-programme characteristics included in the propensity score. Endogenous sorting of students across groups is addressed by using *sezione* fixed effects, building upon the practice that teachers, or the large majority of them, are always assigned to the same *sezione* depending on its prestige.

In Table 3.7 I report the descriptive statistics for PQM schools and for the schools chosen as control. The average of the various dimensions considered is similar - see columns (1) and (2) - and in fact it is not statistical different between the two groups - see column (3). In column (4) I report the estimates of a logistic regression for the probability of being a PQM school in the working sample. It follows that, after the matching, none of the variables included is a good predictor for being a PQM school. This is suggestive of the fact that the matched pair comparison was successful in choosing a group of schools with similar observable characteristics. This, together with the fact that I will use school fixed effects, corroborates the validity of our identification strategy.

Table 3.8 presents descriptive statistics of average students' characteristics for the pre programme cohort among the two groups of schools. The table shows that there are just minor differences between the two groups in terms of average pupils characteristics: pupils in control schools have an higher percentage of mothers working and of highly educated parents.

We decided to divide the schools into three groups, according to average performance in mathematics test score during the pre-intervention year. The aim of this division is to stratify schools according to their socio-economic background. The summary statistics presented in Table 3.9 suggest that the stratification adopted indeed resembles division according to socio economic background. Schools in the bottom tertile are worse than the schools in the other tertiles along various dimensions: the proportion of disable, dropping out, failing and repeating students is much higher in this group. As for student characteristics, we notice that students attending schools in the bottom tertile come from less-advantaged family backgrounds: less mothers are working, less fathers have a high occupational status, the proportion of parents with low education is much higher, and the proportion of parents with high education much lower; the indicator for home possession (HOME) is lower.[3] Therefore, dividing the schools in the groups based on performances in the pre-treatment year, indirectly stratifies for socio-economic and family background, grouping

[3]The variables used to calculate this indicator are: child has a quiet place to study; child has a desk to do his homework; child as a single room for him-self; number of books in the house; house has an internet connection; house has a burglar alarm; house has more than one bathroom; parents have more than one car. Higher values of the score denote better off households.

Table 3.7: Descriptives at the school level of PQM and control schools

	(1) PQM	(2) Control	(3) Difference	(4) Score
Mathematics, percentage of correct answers	0.480	0.489	-0.009	1.418
			(0.010)	(2.861)
Language, percentage of correct answers	0.572	0.584	-0.012	-4.010
			(0.008)	(3.639)
Proportion of permanent teachers	0.892	0.904	-0.012	-0.341
			(0.012)	(1.612)
Student-teacher ratio	9.632	9.931	-0.299	-0.098
			(0.275)	(0.099)
Number of students in the school	402.8	398.4	4.470	0.001
			(26.34)	(0.001)
Proportion of immigrant students	0.027	0.027	0.000	1.088
			(0.003)	(5.101)
Proportion of disable students	0.034	0.031	0.003	2.652
			(0.003)	(8.669)
School drop out rate	0.003	0.003	0.000	-5.368
			(0.001)	(12.07)
School rate of failing students	0.049	0.046	0.003	-3.665
			(0.006)	(3.912)
School rate of repeating students	0.048	0.041	0.007	4.114
			(0.006)	(4.181)
Proportion of females in the school	0.490	0.488	0.002	0.589
			(0.007)	(2.385)
Proportion of classes doing more than 30 hours	0.335	0.337	-0.001	-0.170
			(0.051)	(0.400)
School received PON funds	0.963	0.974	-0.011	-0.273
			(0.023)	(0.769)
Municipality located on mountain	0.284	0.246	0.038	0.186
			(0.056)	(0.306)
(Log) population in town	10.38	10.31	0.069	0.008
			(0.192)	(0.110)
Constant				2.526
				(2.560)
Number of schools	134	114		

Presented in the table are descriptive statistics for the schools entering the final working sample obtained as described in Chapter 3. Column (1) refers to schools participating in the programme; column (2) refers to schools non participating; column (3) reports the difference between column (1) and column (2), together with the corresponding standard error (in parenthesis); column (4) reports the results from a logit regression of the indicator for participating schools on the various dimension considered using only the sample of PQM and control schools, together with the corresponding standard error.

Table 3.8: Descriptive of student average characteristics in PQM and control schools

	(1) PQM	(2) Control	(3) Difference
Test score mathematics	-0.018	0.021	-0.039
			(0.055)
Test score language	-0.026	0.032	-0.059
			(0.050)
Percentage of correct answers mathematics	0.484	0.491	-0.008
			(0.010)
Percentage of correct answers language	0.575	0.585	-0.010
			(0.008)
Proportion of female	0.494	0.492	0.002
			(0.008)
Proportion of ahead students	0.026	0.030	-0.004
			(0.004)
Proportion of behind students	0.060	0.054	0.005
			(0.006)
Proportion of foreign students	0.034	0.032	0.002
			(0.004)
Proportion of students whose mother is working	0.362	0.422	-0.060
			(0.021)
Proportion of students whose father's occupation is : unemployed	0.071	0.052	0.018
			(0.009)
Proportion of students whose father's occupation is : blue collar	0.307	0.304	0.003
			(0.019)
Proportion of students whose father's occupation is : white collar	0.425	0.425	0.001
			(0.017)
Proportion of students whose father's occupation is : managerial	0.197	0.218	-0.022
			(0.016)
Proportion of students whose parents have low education	0.459	0.407	0.052
			(0.027)
Proportion of students whose parents have medium education	0.399	0.415	-0.017
			(0.018)
Proportion of students whose parents have high education	0.142	0.177	-0.035
			(0.018)
Average HOME scale coefficient in the class	-0.066	-0.049	-0.017
			(0.032)
Proportion of students living with both parents	0.901	0.892	0.008
			(0.006)
Class weekly hour	31.76	31.42	0.337
			(0.316)
Class size	21.96	22.43	-0.467
			(-0.467)
Parents' education missing variable	0.234	0.293	-0.059
			(0.041)
Father work missing variable	0.219	0.260	-0.041
			(0.040)
Mother work missing variable	0.189	0.228	-0.039
			(0.041)
Number of schools	134	114	

Presented in the table are descriptive statistics for the schools entering the final working sample obtained as described in Chapter 3. Column (1) refers to schools participating in the programme; column (2) refers to schools non participating; column (3) reports the difference between column (1) and column (2), together with the corresponding standard error (in parenthesis).

in the first group children coming from a more disadvantage background and environment. In the analysis I will consider the whole sample of schools and this stratification.

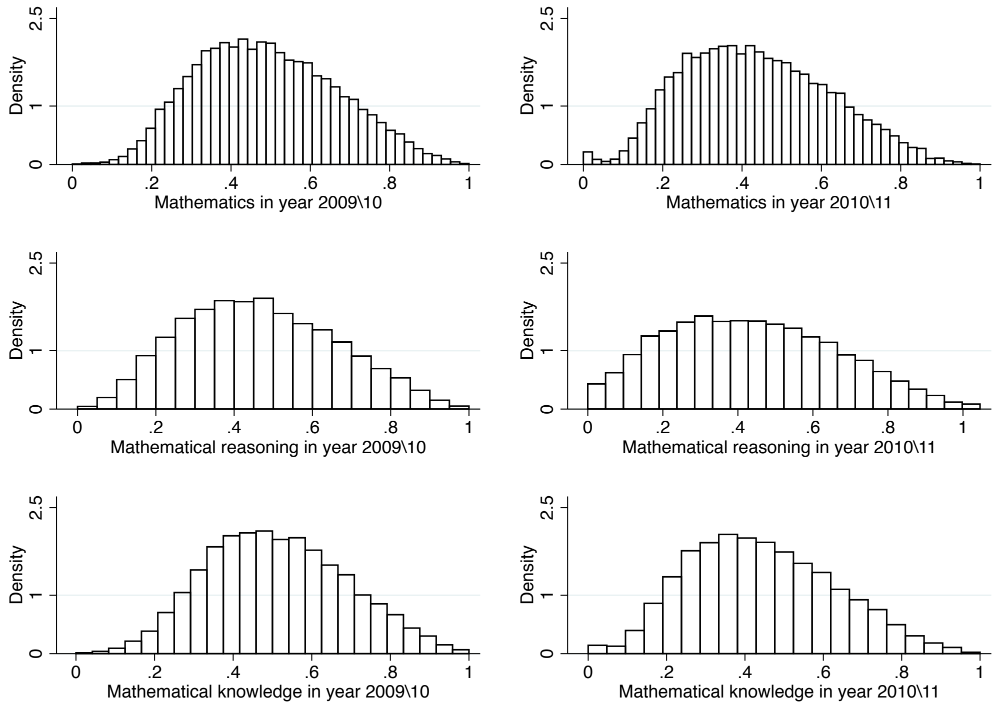
The distribution of the outcome variables that we consider is reported in Figures 3.2, 3.3, 3.4 and 3.5.

Table 3.9: Descriptives in the three groups of schools

	Bottom tertile	Middle tertile	Top tertile
Proportion of permanent teachers	0.874	0.899	0.919
Student-teacher ratio	9.41	10.37	9.52
Number of students in the school	353	463	387
Proportion of immigrant students	0.029	0.022	0.030
Proportion of disable students	0.039	0.029	0.029
School drop out rate	0.006	0.001	0.002
School rate of failing students	0.065	0.037	0.041
School rate of repeating students	0.066	0.039	0.031
Test score mathematics	-0.446	-0.025	0.459
Test score language	-0.323	0.027	0.291
Percentage of correct answers in mathematics	0.405	0.488	0.567
Percentage of correct answers in Italian language	0.530	0.587	0.622
Student is a female	0.492	0.487	0.499
Student one year ahead	0.019	0.029	0.035
Student one year behind	0.081	0.049	0.042
Immigrant student	0.033	0.028	0.037
Child lives with both parents	0.890	0.900	0.901
Parents' education: low	0.526	0.426	0.357
Parents' education: medium	0.366	0.398	0.454
Parents' education: high	0.108	0.176	0.189
Mother is currently working	0.347	0.391	0.430
Father occupation: unemployed	0.075	0.062	0.051
Father occupation: blue collar	0.338	0.303	0.278
Father occupation: clerical white collar	0.434	0.413	0.429
Father occupation: managerial	0.152	0.223	0.243
HOME scale coefficient	-0.170	-0.029	0.022
Class weekly hour	31.59	31.21	31.98
Class size	21.85	22.62	22.09
Number of schools	82	82	84

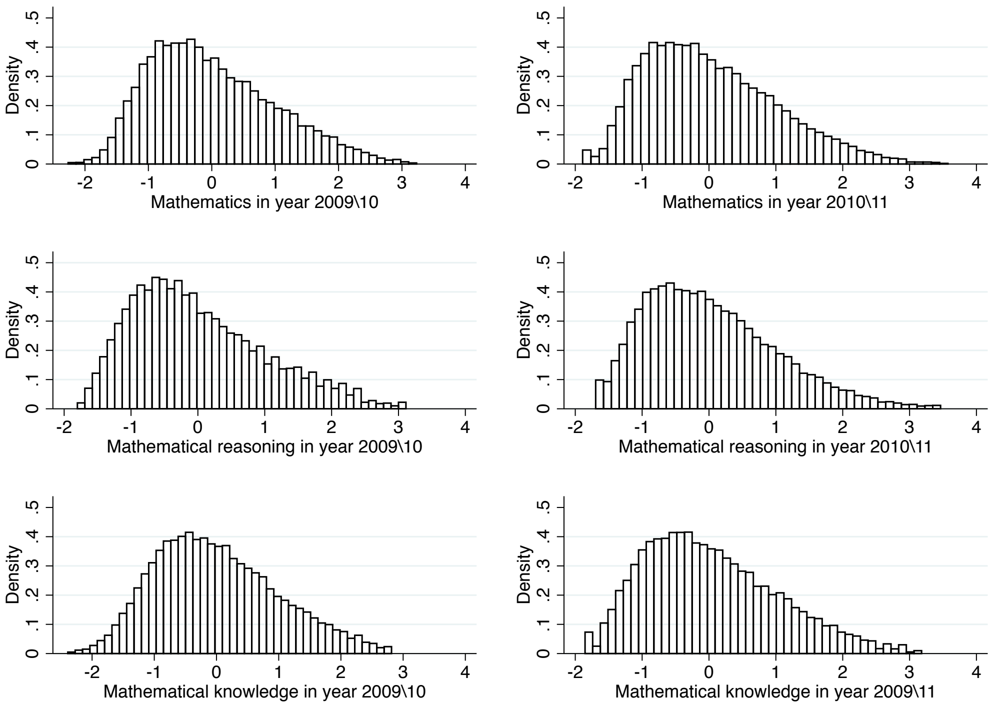
Presented in the table are descriptive statistics for the school entering the final working sample divided in three groups according to average test score in pre-treatment year as described in Chapter 3.1.

Figure 3.2: Distributions of percentage of correct answers in mathematics



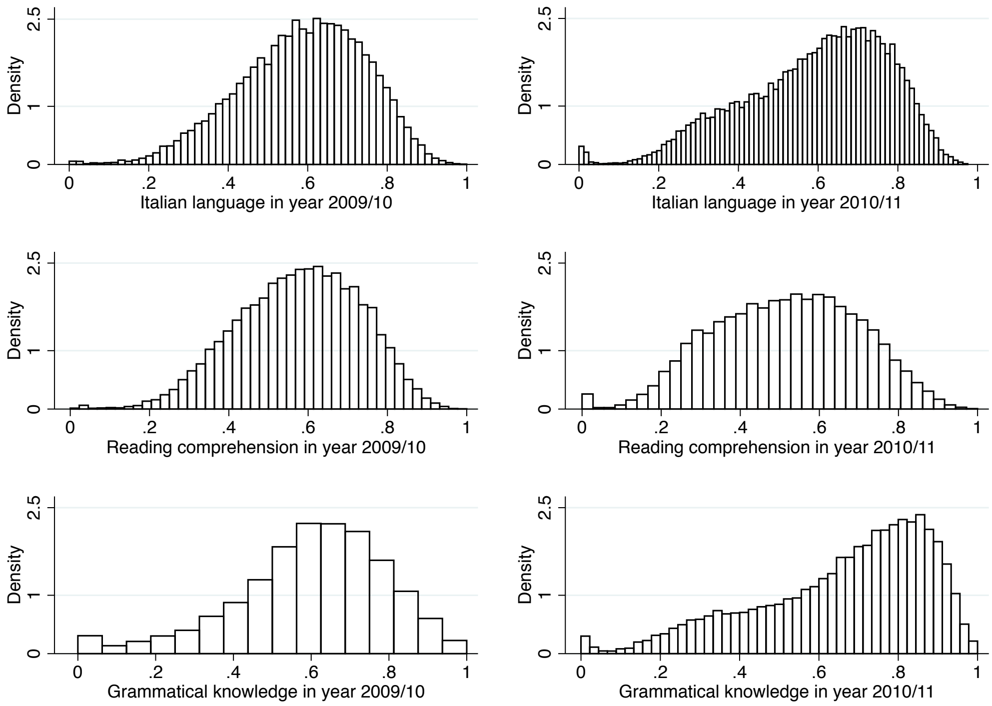
In the Figure the distribution, in the two considered years, in the final working sample of the percentage of correct answers in the whole mathematics test (upper panel); in the mathematical reasoning part of the test (central panel); and in the mathematical knowledge part of the test (bottom panel).

Figure 3.3: Distributions of mathematics test score



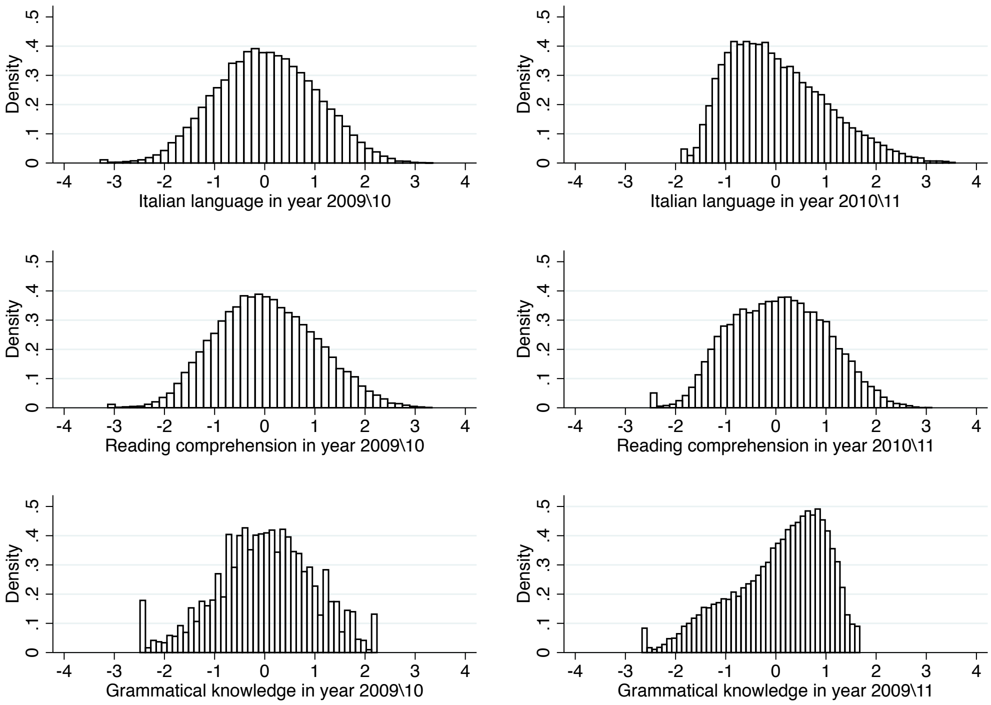
In the Figure the distribution, in the two considered years, in the final working sample of the test score in mathematics (upper panel); in the mathematical reasoning part of the test (central panel); and in the mathematical knowledge part of the test (bottom panel).

Figure 3.4: Distributions percentage of correct answers in Italian Language



In the Figure the distribution, in the two considered years, in the final working sample of the percentage of correct answers in the whole Italian language test (upper panel); in the reading and comprehension part of the test (central panel); and in the grammatical knowledge part of the test (bottom panel).

Figure 3.5: Distributions of Italian language test score



In the Figure the distribution, in the two considered years, in the final working sample of the test score in Italian language(upper panel); in the reading and comprehension part of the test (central panel); and in the grammatical knowledge part of the test (bottom panel).

3.2 Method

Treatment status is defined at the class level and for estimation purposes I use data on the two cohorts of sixth graders in 2009/10 and 2010/11. In practice I compare performances of two contiguous cohorts of children, belonging to the same *sezione* at the end of sixth grade, before and after the programme implementation. This is a standard difference-in-differences approach, with *sezione* fixed effects. The usual assumption needed to use this strategy is that, in the absence of the programme, average change in test scores would have been the same for treated and control groups. Note that, by controlling for *sezione* fixed effect we indirectly control for school fixed effect, and thus for sources of potential biases related to unobservable characteristics of the classes and of the schools.

I'm not able to identify the students who really participate in the afternoon activities, therefore the only effect I can estimate is at the *sezione* level, to be interpreted as the effect of begin in a *sezione* where the teacher have the possibility and the support to organise and provide students with extra instruction time. I will later exploit the number of treated students and the number of activities provided to estimate a treatment intensity, going beyond binary treatment status definition.

Since some classes receive extra education in mathematics, some in language and some in both, I include in the equation both variables, so to control for eventual cross subject effects (i.e effect of PQM mathematics on language outcome and vice-versa), and to take into account that some classes that I use as control in the mathematics (language) equation are actually receiving the treatment in language (mathematics) and thus may be different from real controls (classes not receiving any kind of extra time at school).

My basic specification considers the following equation:

$$y_{jt}^{k} = \alpha + \beta_1 C_{jt}^{M} T_{jt} + \beta_2 C_{jt}^{L} T_{jt} + \beta_3 N_{jt} T_{jt} + \beta_4 T_{jt} + \gamma_j + \theta X_{jt} + \epsilon_{jt}, \tag{3.1}$$

where y_{jt}^{k} is the outcome variable, both percentage of correct answers and test score, in *sezione* j, in year t and in subject k. The treatment effects are estimated using aggregate data at the *sezione* level, considering the *average* result for students in that group. T_{jt} is an indicator for observations in the post-intervention year. C_{jt}^{L} and C_{jt}^{M} are dummies for being enrolled in *any* activity in Italian language (L) and mathematics (M), respectively, while N_{jt} is dummy for control classes in PQM schools. X_{jt} is a set of student and class level variables, γ_j is the *sezione* fixed effect and ϵ_{jt} is a random error. The coefficients β_1 and β_2 are the main effects of interest, while β_3 captures possible spill over effects of treated classes on not treated classes in PQM schools. All standard errors are clustered at the school level.

The basic specification is further refined by considering variability in the number of activities across groups. I construct two measures of treatment intensity. The first exploits variation in the number of activities, and is defined as total number of PQM hours in subject k divided by the total number of hours dedicated to subject k during regular school time (we call this *percentage change in instruction time* for subject k).[4] The second proxy of treatment exposure exploit also variation in the number of students participating to each activity, and it is defined as average number of students in *sezione* j participating in PQM for subject k divided by class size (i.e. the proportion of students actually treated), multiplied by the first measure of intensity. This is the *percentage change in instruction time per student.*

Number of activities[5] and number of students participating are chosen by the teachers, thus these variables are endogenous and can be given a causal interpretation only if we assume that teacher decision about how many activities to propose and how many students to select is a reflection of his personal choice, and thus captured by the class fixed effect. If, on the other side, this decision it is linked to other things which we can not observe and can not be captured with the class fixed effect, that this estimate is not to be considered as causal.

Descriptives of the two measures of intensity in the three groups of schools are provided in Table 3.10. The empirical model for the two measures of intensity is then estimated as follows:

$$y^k_{jt} = \alpha + \beta_1 INT^M_{jt} T_{jt} + \beta_2 INT^L_{jt} T_{jt} + \beta_3 N_{jt} T_{jt} + \beta_4 T_j + \gamma_j + \theta X_{jt} + \epsilon_{jt}, \quad (3.2)$$

where INT^M_{jt} and INT^L_{jt} refer to Italian language (L) and mathematics (M), respectively.

Finally I replicate all the above estimations considering the division into *knowing* and *reasoning* for the mathematics outcomes, and the division into *grammar* and *reading comprehension* for the Italian language outcomes.

As a second step, I also run the same model using micro data at the student level, the corresponding equation is:

$$y^k_{ijt} = \alpha + \beta_1 C^M_{jt} T_{jt} + \beta_2 C^L_{jt} T_{jt} + \beta_3 N_{jt} T_{jt} + \beta_4 T_j + \gamma_j + \theta X_{ijt} + \epsilon_{ijt}, \quad (3.3)$$

where y^k_{ijt} is outcome in mathematics or language of student i in class j, and X_{ijt} is a vector of student and class characteristics.

Estimates at the student level do not add much to the estimates at the class level, nevertheless student level data could be exploit to assess whether there

[4] I know that each activity lasts 15 hours, and I know that children in lower secondary schools dedicate 4 hours per week to mathematics and 7 hours per week to Italian language.

[5] Activities can be at maximum 8.

exist heterogeneous effect in different subgroups of the students population, for example to estimate the effect of male versus female or of foreign versus native students.

Table 3.10: Intensity measures in mathematics and Italian language in the three groups of schools

		Bottom tertile	Middle tertile	Top Tertile
Mathematics				
	25^{th}	0.188	0.281	0.188
INT1	Mean	0.305	0.380	0.309
	75^{th}	0.375	0.375	0.375
	25^{th}	0.107	0.144	0.134
INT2	Mean	0.176	0.202	0.187
	75^{th}	0.197	0.225	0.201
Italian Language				
	25^{th}	0.161	0.161	0.161
INT1	Mean	0.190	0.219	0.178
	75^{th}	0.214	0.268	0.214
	25^{th}	0.073	0.085	0.074
INT2	Mean	0.104	0.109	0.101
	75^{th}	0.119	0.122	0.114

Presented in the table are descriptive statistics of the two measures of intensity in the three groups of schools. Measures calculated as described in Chapter 3.2.
INT1 defined as the percentage change in instruction time in subject j.
INT2 defined as the percentage change in instruction time per student.

3.3 Empirical results

I present results of the average effect of the intervention on both the percentage of correct answers and the test scores (Table 3.11). We see a positive effect of extra time at school in mathematics, on both outcomes in mathematics, with the intervention increasing by 2.5 % the percentage of correct answers, and by 0.134 standard deviations the test score. On the other side no effect is found of extra time at school in language on language test score or on the percentage of correct answers. There are no spill over effect, since the coefficients associate to being a non treated classes in a PQM school are never significantly different from 0, thus non treated classes in PQM schools are not different from non treated classes in control schools, thus the effect of receiving extra hours of education is limited to the classes that are actually enrolled. In addition there are no cross subject effects, since receiving extra time in mathematics does not have an effect on language outcomes, and the same holds for extra time in language which does not have any effect on mathematic outcomes.

In these estimates, as in all the following ones I include as control variables: class size, number of regular schools hours per week, and some controls at the (average) student level: gender, immigration status, regularity status (whether student is ahed or behind compared to his age), maximum level of education of the parents and mother working status; in addition for each of these variables I include a variable indicating the proportion of students in the class with missing value (excluding class size and class weekly timing which show no missing). While there are very few students with missing information about gender, immigration and regularity status, the percentage of students with missing parents' level of education and mother working status is as high as 30%.

Control variables behave as expected: an higher proportion of females in the class is positively associated to language outcomes, and negatively associated to mathematics outcomes. While proportion of ahead students does not have any impact, an higher proportion of behind (i.e. repeating) students, decreases both outcomes in mathematics and language. An higher level of average parental education is associated to higher outcomes in both subjects, and higher proportion of working mothers is positively associated to language outcomes. Variables at the class level, class size and class weekly hours, do not affect neither mathematics nor language.

Table 3.11: Effect of PQM on the percentage of correct answers and on the test scores in mathematics and Italian language

	Percentage of correct answers		Test scores	
	Mathematics	Language	Mathematics	Language
Any extra class in mathematics	0.025*	0.009	0.134*	0.065
	(0.012)	(0.009)	(0.068)	(0.052)
Any extra class in Language	-0.013	-0.000	-0.074	0.007
	(0.010)	(0.008)	(0.060)	(0.048)
Post-treatment cohort	-0.061***	0.009	-0.016	0.031
	(0.006)	(0.005)	(0.036)	(0.030)
Control class in PQM schools	0.003	-0.004	0.020	-0.022
	(0.009)	(0.007)	(0.049)	(0.042)
Proportion of female in the class	-0.066**	0.037*	-0.379***	0.177
	(0.020)	(0.016)	(0.110)	(0.096)
Proportion of ahead students in the class	0.078	0.061	0.391	0.403
	(0.046)	(0.045)	(0.273)	(0.268)
Proportion of behind students in the class	-0.083*	-0.157***	-0.430*	-0.796***
	(0.035)	(0.039)	(0.197)	(0.221)
Proportion of foreign students in the class	-0.061	-0.072	-0.307	-0.437
	(0.054)	(0.042)	(0.304)	(0.241)
Proportion of students whose parents have medium education	0.059**	0.075***	0.292*	0.390***
	(0.021)	(0.016)	(0.114)	(0.096)
Proportion of students whose parents have high education	0.121***	0.114***	0.657***	0.605***
	(0.023)	(0.021)	(0.125)	(0.120)
Proportion of students whose mother is working	0.023	0.037*	0.099	0.192*
	(0.019)	(0.015)	(0.104)	(0.084)
Class weekly hour	-0.000	-0.001	-0.000	-0.007
	(0.002)	(0.001)	(0.013)	(0.009)
Class size	0.001	0.001	0.006	0.004
	(0.001)	(0.001)	(0.004)	(0.004)
Constant	0.456***	0.524***	-0.164	-0.199
	(0.080)	(0.051)	(0.435)	(0.315)
Observations	2622	2622	2622	2622

Difference-in-differences estimates of the effect of the intervention on the percentage of correct answers and test score. Estimates are at the class level with class fixed effect. Standard errors clustered at the school level in parentheses. Each column refers to a separated regression. Estimated using Equation 3.1. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.12: Effect of PQM on the percentage of correct answers and test scores in the three groups of schools

	Percentage of correct answers					
	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
Any extra class in mathematics	0.046*	0.004	0.012	0.010	0.010	0.007
	(0.021)	(0.016)	(0.018)	(0.014)	(0.016)	(0.017)
Any extra class in Language	0.000	-0.011	-0.014	0.005	-0.055*	-0.001
	(0.016)	(0.014)	(0.015)	(0.014)	(0.023)	(0.019)
Control class in PQM schools	0.011	-0.002	-0.008	-0.007	-0.000	-0.007
	(0.014)	(0.013)	(0.012)	(0.009)	(0.017)	(0.016)
	Test scores					
	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
Any extra class in mathematics	0.257*	0.020	0.095	0.103	0.014	0.013
	(0.111)	(0.085)	(0.105)	(0.077)	(0.100)	(0.096)
Any extra class in Language	0.006	-0.084	-0.066	0.050	-0.324*	-0.030
	(0.087)	(0.075)	(0.085)	(0.081)	(0.140)	(0.105)
Control class in PQM schools	0.072	-0.031	-0.040	-0.031	-0.006	-0.066
	(0.074)	(0.068)	(0.063)	(0.047)	(0.101)	(0.091)
Observations	774	774	1018	1018	830	830

Difference-in-differences estimates of the effect of the intervention on the percentage of correct answers and test scores.
Estimates are at the class level with class fixed effect. Standard errors clustered at the school level in parentheses.
Each column refers to a separated regression. Covariates included but not reported.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Tables 3.12 I report the same results but stratifying schools according to average test score in the pre-intervention year. The positive effect of extra time in mathematics on mathematics outcomes is driven by large average returns to participation only for students in the most problematic schools, that is schools in the lowest tertile of student achievement in the pre-programme period, and this effect is positive for both percentage of correct answers and for test scores. Receiving extra time in language has no effects on language outcomes in any of the three groups, but for schools in the top tertile we find that extra hours tailored around language activities have a negative average effect on test scores in mathematics. We do not observe any spill over effects for the non treated classes in PQM schools.

Thus it seams that the intervention is effective only in the schools characterised by a low socio-economic background and more disadvantage situations. Thus, in what follow I will always maintain this stratification.

I then replicate the analysis distinguishing between the *knowing* and in the *reasoning* parts of the mathematics test and between the *grammar* and *reading and comprehension* parts of the Italian language test. Table 3.13 shows the results for the two mathematics domains: the effect of extra instruction time in mathematics is positive and significant in the first group of schools just for the *reasoning* domain, and this holds both for the percentage of correct answers and for the test score; on the other side the effect on the *knowing* domain is positive, but it is not significant. This finding is interesting since it suggest that extra instruction time in the afternoon does not increase the basic knowledge of the targeted children, but it helps them applying and using the knowledge they have to boost their mathematical reasoning and ability to solve mathematical problems. In addition the negative effect of extra time in language on mathematics in the third group is driven by a negative effect on the *knowing* domain.

As for the two language domains, *grammar* and *reading and comprehension* we find that in none of the domains receiving extra hours of education has any effect. (Table 3.14)

Table 3.13: Effect of PQM on mathematical knowledge and mathematical reasoning

	Mathematical knowledge					
	Bottom tertile		Middle tertile		Top tertile	
	% of correct answers	Test score	% of correct answers	Test score	% of correct answers	Test score
Any extra class in mathematics	0.040	0.223	0.008	0.081	0.017	0.072
	(0.024)	(0.125)	(0.017)	(0.098)	(0.016)	(0.083)
Any extra class in Language	-0.001	0.000	-0.008	-0.022	-0.064**	-0.382**
	(0.017)	(0.087)	(0.015)	(0.079)	(0.023)	(0.128)
Control class in PQM schools	0.002	0.027	-0.004	-0.012	-0.002	-0.025
	(0.014)	(0.072)	(0.011)	(0.059)	(0.017)	(0.086)
	Mathematical reasoning					
	Bottom tertile		Middle tertile		Top tertile	
	% of correct answers	Test score	% of correct answers	Test score	% of correct answers	Test score
Any extra class in mathematics	0.056**	0.256**	0.021	0.097	0.003	-0.037
	(0.020)	(0.091)	(0.021)	(0.104)	(0.021)	(0.118)
Any extra class in Language	0.000	0.011	-0.021	-0.091	-0.046	-0.224
	(0.018)	(0.082)	(0.018)	(0.087)	(0.027)	(0.144)
Control class in PQM schools	0.018	0.100	-0.012	-0.059	0.003	0.011
	(0.017)	(0.075)	(0.014)	(0.066)	(0.020)	(0.109)
Observations	774	774	1018	1018	830	830

Difference-in-differences estimates of the effect of the intervention on the percentage of correct answers and test scores in the mathematical knowledge and mathematical reasoning .
Estimates are at the class level with class fixed effect. Standard errors clustered at the school level in parentheses.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
Each column refers to a separated regression. Covariates included but not reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.14: Effect of PQM on language reading and comprehension and grammar knowledge

	Reading and comprehension					
	Bottom tertile		Middle tertile		Top tertile	
	% of correct answers	Test score	% of correct answers	Test score	% of correct answers	Test score
Any extra class in mathematics	0.005	0.026	0.014	0.122	0.000	-0.012
	(0.014)	(0.079)	(0.013)	(0.078)	(0.016)	(0.096)
Any extra class in Language	-0.011	-0.086	0.006	0.064	-0.007	-0.064
	(0.012)	(0.069)	(0.013)	(0.073)	(0.019)	(0.110)
Control class in PQM schools	-0.006	-0.051	-0.007	-0.027	-0.012	-0.089
	(0.011)	(0.064)	(0.009)	(0.050)	(0.014)	(0.084)
	Grammar knowledge					
	Bottom tertile		Middle tertile		Top tertile	
	% of correct answers	Test score	% of correct answers	Test score	% of correct answers	Test score
Any extra class in mathematics	0.011	0.050	0.003	0.024	0.008	0.006
	(0.022)	(0.099)	(0.019)	(0.092)	(0.019)	(0.089)
Any extra class in Language	-0.003	-0.007	0.001	0.011	0.007	0.042
	(0.019)	(0.084)	(0.019)	(0.090)	(0.021)	(0.100)
Control class in PQM schools	0.003	0.010	-0.010	-0.058	-0.006	-0.049
	(0.016)	(0.073)	(0.011)	(0.047)	(0.020)	(0.093)
Observations	774	774	1018	1018	830	830

Difference-in-differences estimates of the effect of the intervention on the percentage of correct answers and test scores in language reading and comprehension and grammar knowledge.
Estimates are at the class level with class fixed effect. Standard errors clustered at the school level in parentheses.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
Each column refers to a separated regression. Covariates included but not reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As we previously explained, not all the classes are receiving the same amount of treatment, since the number of activities and number of students varies, thus we replicate the analysis to assess the impact of treatment intensity on mathematics and language outcomes. Table 3.15 provides estimates of the average effect of treatment intensity on mathematics and language outcomes, when we define intensity as the percentage increase in instruction time in a given subject. An higher increase in instruction time in mathematics has a positive effect on the mathematics test score and percentage of correct answers in the group of more disadvantage schools; as usual we find no effect on the the other two groups, and no effect of more instruction time in language on language test scores. Table 3.16 provides the same results for intensity defined as the average percentage increase in instruction time per student, thus when we take into account not only the number of activities done but also the number of students involved in each activity. More extra time in mathematics per students has a positive effect on test score in the first group of schools, and we also find the negative impact of more time per student dedicate to extra activities in language on mathematics test score in the top tertile group of schools.

Finally, I want to explore other heterogeneity dimensions exploiting students characteristics, to asses whether the intervention has different effects on different subgroups of the population. In particular I estimate the effect for female vs male and for foreign vs native. Thus I replicate the analysis, using student level data and maintaining the stratification of schools in three groups according to pre-intervention year performances, interacting the dummies receiving extra time in mathematics and language with the dummy for student's gender; then with the dummy for student's origin. I'm running these specifications focusing just on test scores, thus leaving aside the percentage of correct answers.

Table 3.17 present the results. We do not see big differences between males and female: the effect of extra time at schools engaged in mathematics activities has a positive effect on mathematics test score for both males (β = 0.283; se = 0.115) and females (β = 0.295; se = 0.118) just in the first group of schools. The negative effect of more time dedicated to language on mathematics test score found in the third group is the same for both males (β = -0.315; se = 0.137) and females (β = -0.329; se = 0.137). Secondly I also wanted to see whether there were differences between males and females in the different domains, thus I replicate the analysis for the *knowing* and *reasoning* parts in mathematics and for the *grammar* and *reading and comprehension* parts in Italian language. Interestingly we see that while the positive effect in the schools belonging to the lowest tertile, for males is found only in the *reasoning* part ($\beta = 0.295$; se = 0.095), for females we find a positive effects in both dimensions, (*knowing* : β= 0.265; se = 0.129; *reasoning*: $\beta = 0.284$; se = 0.101) (Table 3.18). No effect is found for neither males or females in none of the two Italian language domains.[6]

As a further check I wanted to see whether the intervention could reduce the well known gap in mathematics performances that exist between boys and girls (with boys outperforming girls in mathematics). Therefore I calculate the ratio of percentage of correct answers given by girls over the percentage of correct answers given by boys in the whole mathematics test, and in the knowing and reasoning parts, and then estimate whether this ratio was changed by spending more time at school. In Table 3.19, we notice no effect on the ratio in the first group of schools. Thus the intervention does not significantly discriminate between boys and girls, since both increase their test scores in mathematics, but the programme does not manage to close the gap between girls and boys in mathematics. Interestingly we find that the ratio increases in the top tertile group of schools for the knowing part of the test: thus, although the intervention did not significantly increases the performances of children in this group, it managed to slightly reduce the gap between boys and girls in the knowing dimension.

As for the difference between foreign and native students we see that the positive effect for mathematics found in the first group of school is driven just by native students, while foreign students do not benefit from the programme ($\beta = 0.161$; se = 0.173). (Table 3.20).

[6] Corresponding Table not included.

Table 3.15: Effect of treatment intensity on the percentage of correct answers and test scores

	Percentage of correct answers					
	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
INT1 mathematics	0.121*	-0.009	0.018	0.000	0.076	0.083
	(0.060)	(0.050)	(0.046)	(0.030)	(0.051)	(0.056)
INT1 language	-0.020	-0.094	-0.092	-0.003	-0.238	0.101
	(0.064)	(0.069)	(0.046)	(0.063)	(0.120)	(0.104)
Control class in PQM schools	0.007	-0.005	-0.010	-0.009	0.004	-0.001
	(0.013)	(0.012)	(0.011)	(0.009)	(0.017)	(0.016)
	Test scores					
	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
INT1 mathematics	0.687*	-0.050	0.167	0.138	0.234	0.322
	(0.310)	(0.269)	(0.270)	(0.169)	(0.296)	(0.330)
INT1 language	-0.050	-0.607	-0.476	0.096	-1.446*	0.432
	(0.339)	(0.367)	(0.257)	(0.338)	(0.704)	(0.585)
Control class in PQM schools	0.055	-0.049	-0.052	-0.045	0.016	-0.036
	(0.068)	(0.064)	(0.060)	(0.046)	(0.099)	(0.090)
Observations	774	774	1018	1018	830	830

Difference-in-differences estimates of the effect of the treatment intensity on the percentage of correct answers and test scores.
Intensity defined as the percentage change in instruction time in subject j.
Estimates are at the class level with class fixed effect. Standard errors clustered at the school level in parentheses.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
Each column refers to a separated regression. Covariates included but not reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.16: Effect of treatment intensity on the percentage of correct answers and test scores

	Percentage of correct answers					
	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
INT2 mathematics	0.171	-0.038	0.003	0.025	0.109	0.101
	(0.089)	(0.082)	(0.082)	(0.051)	(0.084)	(0.095)
INT2 language	-0.085	-0.167	-0.148	0.029	-0.468*	0.129
	(0.133)	(0.107)	(0.110)	(0.114)	(0.207)	(0.202)
Control class in PQM schools	0.004	-0.007	-0.010	-0.008	0.003	-0.003
	(0.013)	(0.011)	(0.011)	(0.008)	(0.017)	(0.015)
	Test scores					
	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
INT2 mathematics	1.012*	-0.206	0.099	0.388	0.271	0.380
	(0.467)	(0.445)	(0.466)	(0.291)	(0.472)	(0.536)
INT2 language	-0.337	-1.099	-0.834	0.382	-2.836*	0.568
	(0.713)	(0.579)	(0.597)	(0.639)	(1.168)	(1.099)
Control class in PQM schools	0.039	-0.056	-0.057	-0.037	0.009	-0.044
	(0.069)	(0.062)	(0.060)	(0.045)	(0.098)	(0.089)
Observations	774	774	1018	1018	830	830

Difference-in-differences estimates of the effect of the treatment intensity on the percentage of correct answers and test scores.
Intensity defined as the percentage change in instruction time in subject per student.
Estimates are at the class level with class fixed effect. Standard errors clustered at the school level in parentheses.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1.
Each column refers to a separated regression. Covariates included but not reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.17: Effect of PQM on the test score of males and females

	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
Any extra class in mathematics	0.283*	0.049	0.079	0.076	-0.003	-0.007
	(0.115)	(0.103)	(0.106)	(0.087)	(0.104)	(0.098)
Any extra class in mathematics * female	0.011	-0.001	-0.004	0.040	0.041	0.037
	(0.063)	(0.063)	(0.072)	(0.057)	(0.069)	(0.059)
Any extra class in Language	0.046	-0.082	-0.121	-0.043	-0.315*	-0.007
	(0.087)	(0.085)	(0.087)	(0.089)	(0.137)	(0.102)
Any extra class in language * female	-0.010	0.106	0.087	0.141**	-0.014	-0.014
	(0.051)	(0.054)	(0.060)	(0.047)	(0.061)	(0.059)
Control class in PQM school	0.068	-0.034	-0.068	-0.043	-0.039	-0.088
	(0.066)	(0.065)	(0.065)	(0.050)	(0.104)	(0.092)
Observations	15060	15060	21607	21607	17627	17627

Difference-in-differences estimates of the effect of the intervention on the test scores.
Estimates are at the student level with class fixed effect. Standard errors clustered at the school level in parentheses.
Each column refers to a separated regression. Covariates included but not reported.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.18: Effect of PQM on the test score of males and females in the mathematics knowing and reasoning dimensions

	Bottom tertile		Middle tertile		Top tertile	
	Knowing	Reasoning	Knowing	Reasoning	Knowing	Reasoning
Any extra class in mathematics	0.232	0.295**	0.079	0.074	0.015	-0.018
	(0.128)	(0.095)	(0.103)	(0.103)	(0.093)	(0.121)
Any extra class in mathematics * female	0.033	-0.010	-0.019	0.005	0.112	-0.023
	(0.060)	(0.064)	(0.073)	(0.066)	(0.073)	(0.070)
Any extra class in Language	0.026	0.059	-0.076	-0.139	-0.372**	-0.218
	(0.092)	(0.080)	(0.085)	(0.088)	(0.131)	(0.139)
Any extra class in language * female	-0.003	-0.016	0.104	0.059	0.013	-0.036
	(0.056)	(0.049)	(0.063)	(0.059)	(0.068)	(0.064)
Control class in PQM school	0.018	0.100	-0.037	-0.085	-0.051	-0.023
	(0.066)	(0.068)	(0.059)	(0.067)	(0.091)	(0.110)
Observations	15060	15060	21607	21607	17627	17627

Difference-in-differences estimates of the effect of the intervention on the mathematics knowing and reasoning test scores.
Estimates are at the student level with class fixed effect. Standard errors clustered at the school level in parentheses.
Each column refers to a separated regression. Covariates included but not reported.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.19: Effect of PQM on the ratio of correct answers given by females and males

	Bottom tertile			Middle tertile			Top tertile		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Math	Knowing	Reasoning	Math	Knowing	Reasoning	Math	Knowing	Reasoning
Any extra class in mathematics	-0.011	-0.019	-0.022	0.006	-0.001	-0.017	0.070	0.122*	0.038
	(0.045)	(0.056)	(0.065)	(0.042)	(0.052)	(0.059)	(0.045)	(0.061)	(0.061)
Any extra class in language	0.035	0.035	0.101	0.066	0.087	0.035	0.030	0.035	0.052
	(0.042)	(0.058)	(0.063)	(0.051)	(0.060)	(0.071)	(0.038)	(0.047)	(0.056)
Control class in PQM school	-0.018	-0.003	-0.023	-0.039	-0.032	-0.082*	0.045	0.048	0.056
	(0.031)	(0.042)	(0.049)	(0.025)	(0.028)	(0.034)	(0.035)	(0.042)	(0.050)
Observations	773	773	773	1017	1017	1017	830	830	830

Difference-in-differences estimates of the effect of PQM on the ratio of percentage of correct answers given by females over the percentage of correct answers given by males in the whole mathematics test (columns (1), (4) and (7)); in the knowing part of the test (columns (2), (5) and (8)) and in the reasoning part of the test (columns (3), (6) and (9)).

Estimates at the class level with class fixed effect. Standard errors clustered at the school level in parenthesis

Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.20: Effect of PQM on the test score on native and foreign

	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
Any extra class in mathematics	0.300**	0.041	0.065	0.087	-0.012	-0.039
	(0.113)	(0.091)	(0.109)	(0.081)	(0.104)	(0.097)
Any extra class in mathematics * foreign	-0.140	-0.052	-0.251	0.033	-0.261	-0.138
	(0.111)	(0.116)	(0.159)	(0.183)	(0.233)	(0.252)
Any extra class in Language	0.036	-0.034	-0.090	0.014	-0.319*	-0.004
	(0.085)	(0.075)	(0.084)	(0.079)	(0.143)	(0.103)
Any extra class in language * foreign	0.089	0.132	0.067	0.030	0.278	0.169
	(0.159)	(0.158)	(0.149)	(0.173)	(0.206)	(0.142)
Control class in PQM school	0.068	-0.034	-0.064	-0.041	-0.034	-0.081
	(0.066)	(0.065)	(0.064)	(0.050)	(0.103)	(0.092)
Observations	15060	15060	21607	21607	17627	17627

Difference-in-differences estimates of the effect of the intervention on the test scores.
Estimates are at the student level with class fixed effect. Standard errors clustered at the school level in parentheses.
Each column refers to a separated regression. Covariates included but not reported.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.4 Robustness check

It is worthwhile provide some robustness checks of the results I find with my main model specification.

The first thing I want to test is whether future enrolled in PQM programme has any effect of pre-programme test scores. In practice I run a regression using only pre-programme data at the class level, in which the main dependent variable is begin a *sezione* which next year will be enrolled in the PQM programme. This regression will help in establishing whether the *sezioni* chosen to participate in the programme were different, in terms of test scores, from other the *sezioni* of the school. The equation includes the usual variables at the class and student level, school fixed effect, and standard errors are clustered at the school level. In the specification used I included only observations belonging to PQM schools, i.e. excluding classes in control schools.

Table 3.21 shows that future participation on the language programme had a positive effect on language and mathematics test score in the middle and top tertile groups, while had a positive effect on language test scores just in the top tertile group. On the other side future participation on the mathematics programme had no effect on any test scores in any group. These results suggest that the classes chosen to participate in the language programme, in the two top tertile groups, were the classes that were already performing better in language and mathematics, compare to other classes in the school; while this is not true for the classes chosen to participate in the math programme, which are not different from the other classes.

Given the specification I am using to control for sorting of teachers and students (i.e.*sezione* fixed effect), it s not really an issue whether classes chosen to participate to the programme were different from the classes not chosen, nevertheless it is reassuring that I do not find evidence of hugh selection of classes in teachers int he mathematics part of the programme.

As a second check I want to exclude the possibility that the intervention is acting on other inputs of the achievement production function, and that the effect found it is not the direct effect of the programme, but the effect of other inputs that have been changed due to the programme. In order to exclude this possibility I tested whether the intervention had any effect on any of the other inputs we control for (i.e. students characteristics, class size and number of weekly hour). Thus I run several regressions in which the independent variables are the inputs. More in details, for each of the inputs, I estimated the following equation:

$$x_{ijt} = \alpha + \beta_1 C^M_{jt} T_j + \beta_2 C^L_{jt} T_j + \beta_3 N_{jt} T_j + \beta_4 T_j + \gamma_j + \epsilon_{ijt},$$

where x_{ijt} is the input for student i in class j at time t and the other variables have the same meaning as in equation 3.1. I rely on student level data to better account for the missing value issue: thus in each of these regressions I excluded students who report a missing information for the input considered, but I also estimate the effect of the intervention on the dummies capturing whether a given variable was missing. Table 3.22 shows that the programme did not have any effect on any of the inputs considered, thus the effect found is to consider the direct effect of the intervention and not the indirect effect of changing other inputs that may affect achievement.[7]

[7] Only in the group of school belonging to the second tertile, the intervention increases the number of ahead students.

Table 3.21: Effect of future PQM participation on the test scores

	Bottom tertile		Middle tertile		Top tertile	
	Mathematics	Language	Mathematics	Language	Mathematics	Language
Future participation in PQM mathematics	-0.002	-0.020	0.118	-0.094	-0.016	-0.036
	(0.059)	(0.053)	(0.079)	(0.055)	(0.095)	(0.080)
Future participation in PQM language	-0.002	0.050	0.162*	0.018	0.266*	0.211**
	(0.053)	(0.048)	(0.075)	(0.063)	(0.131)	(0.073)
Observations	263	263	257	257	196	196

Estimates at the class level with school fixed effects. Standard errors clustered at the school level in parenthesis.
Estimated using just pre-intervention year, just in treated schools.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.22: Effect of PQM on the inputs of the achievement production function

	Bottom tertile		Middle tertile		Top tertile	
	Extra time in Mathematics	Extra time in Language	Extra time in Mathematics	Extra time in Language	Extra time in Mathematics	Extra time in Language
Student is a female	-0.014	-0.008	0.002	0.002	-0.021	-0.010
	(0.026)	(0.027)	(0.023)	(0.024)	(0.022)	(0.025)
Ahead student	-0.009	0.009	0.018*	0.000	0.011	0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)
Behind student	-0.009	-0.019	0.013	0.012	0.004	-0.008
	(0.016)	(0.015)	(0.013)	(0.010)	(0.010)	(0.009)
Student regularity status missing	-0.003	-0.005	0.055	0.006	0.002	0.010
	(0.009)	(0.010)	(0.003)	(0.004)	(0.006)	(0.003)
Foreign student	0.003	-0.004	-0.004	0.004	-0.013	0.001
	(0.008)	(0.009)	(0.008)	(0.009)	(0.010)	(0.009)
Student origin missing	-0.020	-0.051	-0.067	-0.043	0.061	-0.039
	(0.050)	(0.050)	(0.046)	(0.037)	(0.041)	(0.019)
Parents' level of education low	-0.064	-0.053	0.005	0.011	0.061	-0.009
	(0.034)	(0.035)	(0.033)	(0.029)	(0.030)	(0.030)
Parents' level of education medium	0.017	0.028	-0.027	-0.017	-0.032	0.001
	(0.032)	(0.033)	(0.032)	(0.022)	(0.030)	(0.035)
Parents' level of education high	0.047	0.024	0.021	0.006	-0.030	0.008
	(0.025)	(0.023)	(0.024)	(0.019)	(0.023)	(0.022)
Parents' level of education missing	0.035	0.069	-0.029	-0.051	0.014	-0.138
	(0.102)	(0.094)	(0.075)	(0.063)	(0.064)	(0.082)
Mother is working	0.055	0.009	-0.019	-0.024	-0.023	0.013
	(0.032)	(0.030)	(0.026)	(0.025)	0.036	0.032
Mother's working status missing	0.091	0.111	-0.074	-0.083	0.032	-0.164
	(0.110)	(0.102)	(0.077)	(0.065)	(0.061)	(0.086)
Class weekly hours	0.180	0.063	-0.271	-0.0494	-0.195	-0.293
	(0.232)	(0.257)	(0.248)	(0.266)	(0.291)	(0.244)
Class size	1.101	-0.048	-0.418	0.003	0.394	1.043
	(0.668)	(0.601)	(0.626)	(0.647)	(0.729)	(0.587)

Difference-in-differences estimates of the effect of the intervention on the main inputs.
Estimates are at the student level with class fixed effect. Standard errors clustered at the school level in parentheses.
Schools have been divided into three groups according to test scores in pre-treatment year, see Paragraph 3.1
Each row refers to three separated regressions, one for each tertile group. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5 Discussion and conclusion

The aim of this chapter is to provide evidence on the effectiveness of a programme providing students in low achieving schools with extra instruction time in mathematics and Italian language.

The PQM programme increases the number of hours children spend at school, in particular increasing either the mathematics and/or Italian language hours in a selected number of classes of the schools chosen to participate. Throughout a matching, we select a group of control schools which share similar observable characteristics with the chosen one, and using repeated observation of the same *sezione* over time, we are able to take into account sorting of children and teachers in the selected classes.

Results show that extra time at school spent doing mathematics activities increases the test score and the percentage of correct answers in mathematics, just in schools characterised by lower pre-intervention performances (schools whose average performances in the pre-treatment year where below the first tertile of the distribution of test scores), while no effect is found for treated classes in schools belonging to the other two tertiles. This result is in line with the recent study by Lavy (2012), which also find that a similar intervention was more effecting for students coming from low socio-economic background.

Indeed, I believe that children belonging to the first tertile group are the ones coming from the lower socio-economic background, and we interpret the positive effect found in mathematics in two ways: the extra time they spend during the PQM programme is the only time they dedicate to studying outside regular school time, thus they are actually spending more time on academic activities, which means that achievement works as a cumulative process, and more time at schools results in higher performances. Moreover spending more time at school also decreases the "negative" influence of the families, which I assume are not very supportive and helpful for the children in contexts characterised by low socio-economic background. Thus, children involved in the intervention spend at school the time they would otherwise spend doing nothing at home. On the other side, I think that in environments with higher socio-economic background, the PQM programme is working as a substitute for the work the children would anyway do at home and so it is not effective, since also without the programme, the students would probably dedicate some time to academic activities during the afternoon.

In addition it is interesting that the effect is found just for mathematics, while nothing it is found for reading skills. Nevertheless this results it is not new in the literature (see for example Sims (2008)), since it has been shown that is much harder to intervene on reading and comprehension skills, rather than on skills involving a lot of exercise and practice, such as mathematics.

In addition we estimate the effect on the two dimensions captured by the test used: a *knowing* dimension, measuring the knowledge of mathematical contents, and an *reasoning* dimension measuring the ability to use mathematical concepts to solve problems and processes. We find that the effect is positive and significant just on the *reasoning* dimension, meaning that extra instruction time hold outside regular school time does not add much in terms of concepts, but it may be very useful in helping students making the most out of their knowledge, thus stimulating and boosting their ability to think, use and apply their knowledge.

There are no difference in returns for boys versus girls, since both seem to benefit in the same way, thus the intervention does not help in reducing the gap between gigls and boys in mathematics performances. Moreover the effect is driven just by native Italian students, with no effect for foreign.

Finally we investigate whether different exposure to the treatment had any different effect on mathematics outcome, thus we define two measures of treatment intensity, exploiting variability in the number of activities done in the afternoon and in the number of students involved in each activity. We define the first measure as percentage increase in instruction time in subject j and the the second measure as average percentage increase in instruction time per student. Using both measures we find a positive effect on mathematics test score in the first group of schools, thus the more the children are exposed, the more they seem to learn.

Extra time spend at school doing Italian language activities has no effect on language outcomes, and has a negative effect on mathematics test score in the top tertile group of schools. This result suggests that additional time spent at school in reading activities may substitute the time the children would have otherwise invested in studying mathematics. This results, found just in the top tertile, may confirm the conclusion I was drawing before about PQM working as a substitution for the time children would have spent at home. These children are spending time at schools engaged in Italian language activities, and this time substitute in part, the time they would have spent at home maybe practicing mathematics. This may be confirmed also by the fact that the higher the treatment intensity in language, the higher is the loss in mathematics outcomes.

Chapter 4

Non linear effects of extra instruction time on student achievement

The focus of the previous chapter has been on average treatment effects of the intervention on test scores. Nevertheless average treatment effects tell only a small part of story, indeed although we are aware that the program, on average, increases test scores in treated classes, we do not know, for example, whether all the classes are benefiting from the intervention, or whether all the classes are benefiting in the same amount. Similarly we may be interested in finding which students contribute to the average positive effects: is everybody increasing his test scores, or are just the worse (best) students receiving a significant return? In other words, difference-in-differences can reveals whether an intervention was "good" or "bad" on average, but we miss information on single classes and students responses.

This may be a critical point, especially if the aim of the evaluation is to suggest policy implication. Therefore I use non-linear difference-in-differences methods to investigate the effects across *sezione* of participating and non-participating schools, and across students within *sezione* involved in the programme. Non linear methods allow me to go beyond mean impacts, and I can answer questions about the heterogeneity of the effect across the distribution.

The first source of variability that I consider is that coming from outcome differences between classes. As the parameter retrieved through a standard difference-in-differences is the average effect on class test scores, I want to investigate the extent of variability in the returns to participation across classes. Second, I want to study if the possible effects of PQM on class test scores are driven by returns that are markedly different amongst students in the class, thus shedding light on within class variability. Between class variability focuses on distribution of average classes test score (or percentage of correct answers), and it suggests whether all the classes benefit (lose) in the same way from the intervention; within class variability focuses on the distribution of test score (or percentage of correct answers) at the student level and it suggests whether all the students benefit (lose) in the same way.

I rely on two different non linear difference-in-differences methods. The first one is proposed by Firpo et al. (2009) and further extended to the difference-in-differences setting by Havnes and Mogstad (2010); while the second one is proposed by Athey and Imbens (2006). Although the two methods rely on different assumptions, they lead to the similar results, which make me confident about my findings.

The method proposed by Firpo et al. (2009) is used to estimate unconditional quantile treatment effect, and it was expanded by Havnes and Mogstad (2010) to the difference-in-differences framework. They call this method "threshold difference-in-differences", since their idea is to estimate the probability that a class (or a student) test score is above a given threshold, as a way to estimate the effect on unconditional quantiles of the outcome distribution. This approach provide a convenient setting to account for the availability of multiple control groups, and to model treatment intensity represented by variability across PQM classes in the number of activities and students involved. The method proposed by Athey and Imbens (2006), called the "change-in-changes" model, is an extension of the standard difference-in-differences model, and it uses the distributions of the outcome in the three observable non treated groups (non treated pre and post and treated pre) to recover the entire counterfactual distribution of the outcome in the treated group in the post program period had it not been treated. Both methods are quite useful since it they are very simple to estimate, but the first one allows to easily include covariates in the model (which is slightly more complicated in the model proposed by Athey and Imbens (2006)).

It is worth bearing in mind that both methods identify the effects on the distribution of the outcomes, which is different from the distribution of effects on the outcomes,unless the effects are rank-preserving (Heckman et al., 1997)

Before explaining in details the two methodologies, let me introduce some common notation. Let me assume for simplicity that there are just four groups: treated, pre and post and control, pre and post.

Let me define $F_{Y,DT}$ as the cumulative distribution function of outcome y for group $D = d$ in period $T = t$ (which in our case is the classes' average or the students' test scores distribution), where $D = 1$ indicates the treatment group and $T = 1$ indicates post intervention year. In addition let Y^I be the outcome if treated, which is observed if $D = 1, T = 1$ and Y^N be the outcome if untreated, which is observed for the three remaining groups. The aim is to find the counterfactual distribution in the group $D = 1, T = 1$, which is defined as $F_{Y^N,11}(y)$, useful to recover effects on the outcome distribution.

The two quantities of interested are: the intervention effect at a particular level of outcome y, denoted by $\tau(y)$:

$$\tau(y) = F_{Y^I,11}(y) - F_{Y^N,11}(y),$$

and the quantile treatment effects, denoted by $\tau(q)$

$$\tau(q) = F^{-1}_{Y^I,11}(q) - F^{-1}_{Y^N,11}(q).$$

By knowledge of one of the two, the other can be easily recovered.

4.1 The threshold difference-in-differences

The threshold difference-in-differences estimator is defined by Havnes and Mogstad (2010), and it recovers directly $\tau(y)$, by estimating the probability that a class or student test score is above a given level of y. I thus estimate the same regressions in (3.1) and (3.2), where the outcome variable is now a dummy for scoring above pre-specified thresholds, which I set to be subject specific. Thresholds are defined using baseline data, calculating a grid from the 1^{st} to the 99^{th} percentile of the relevant score distribution.

I first consider between *sezione* variability in test scores, and define y^k_{jt} as an indicator for having the average test score in *sezione* j above a certain threshold:

$$y^k_{jt} = \beta_0 + \beta_1 C^M_{jt} T_{jt} + \beta_2 C^L_{jt} T_{jt} + \beta_3 N_{jt} T_{jt} + \beta_4 T_{jt} + \delta_j + \theta X_{jt} + \epsilon_{jt},$$

where all the variables have the same meaning as in the standard difference-in-differences model presented in the previous chapter.

I then consider within class variability in test scores, and define outcomes at the student level. In particular, I estimate the following regression model:

$$y^k_{ijt} = \delta_0 + \delta_1 C^M_{jt} T_{jt} + \delta_2 C^L_{jt} T_{jt} + \delta_3 N_{jt} T_{jt} + \delta_4 T_{jt} + \delta_j + \theta X_{jt} + \epsilon_{jt},$$

where the outcome y^k_{ijt} is defined as a student-specific dummy for scoring above a certain threshold.

Thus, for each value of y the estimate of $\beta_1, \beta_2, \delta_1, \delta_2$ give the effect of the intervention on the probability of scoring at least y in subject k.

The underlying assumption is that the outcome growth from pre-treatment to post-treatment year around each particular level of outcome y is the same in the two groups. I rely throughout on linear probability equations. Form knowledge of the effects at different values of y I can easily back out quantile treatment effects, $\tau(q)$.

4.2 The change-in-changes

As a further check I also implement the methods developed by Athey and Imbens (2006), which they call the "change-in-changes" (CIC). This method relies on slightly different assumptions and it recovers directly the entire counterfactual distribution of the outcome of the treated group had they not been treated. For the application of this method I consider just within class variability, hence I focus on student level data and on distribution of test scores at the student level.[1]

They assume that in absence of the treatment, the outcomes satisfy:

$$Y_i^N = h(U_i, T_i),$$

where $h(u,t)$ is increasing and strictly monotonic in u and doesn't vary with group. U_i is a random variable that represents the unobservable characteristics of unit i, and can vary between groups, but it is constant within groups over time ($U \perp T|D$). Hence $h(u,t)$ is a production function which is equal for all the units, and the outcome of an unit with $U = u$ will be the same in a given time period, irrespective of the group membership. In my framework we can think about U has either ability of the students or also as ability of the teachers: it doesn't matter whether U is different between the two groups, as long as in each group it is the same in both periods. This assumption is coherent with the assumption I made throughout all the thesis, considering repeated observation of the same *sezione* over time as a solution for sorting of students and teachers in the different classes. The monotonicity of h assumption is plausible since higher ability students normally have higher test scores. A further assumption is that the support of U in group 1 is contained in the support of U in group 0, which again it is plausible in my setting, especially since I have as control group a set of schools which have been chosen to be as similar as possible to the PQM schools, hence I expect that I find students and teachers with similar unobserved characteristics.[2]. Finally the CIC accommodates the possibility of selection into treatment due to larger expected benefits from treatment, since it consider the possibility that treated individuals may benefit more from the treatment than untreated individuals and that the intervention may have been implemented because larger gains were expected for the treatment group.

[1] This choice is mainly due to the fact that the method heavily relies on distributions, and considering the data at the class level, which had to be divided into three groups of schools, and furtherly divided into four groups of treated-control, pre-post, we do not have enough observations to fully recover the average distribution of class test scores.

[2] Athey and Imbens (2006) shows that their main findings hold also in absence of common support.

The core idea of the CIC is that the entire distribution of outcomes for the treatment group would experience the same changes over time as the distribution of outcomes for the control group in the absence of the intervention.
Athey and Imbens (2006) show that the distribution of Y_{11}^N is identified and given by:

$$F_{Y^N,11}(y) = F_{Y,10}(F_{Y,00}^{-1}(F_{Y,01}(y)))$$

With the entire counterfactual distribution of outcomes available, it easy to recover both the average treatment effects:

$$\tau^{CIC} = E[Y_{11}^I] - E[F_{Y,01}^{-1}(F_{Y,00}(Y_{10}))],$$

and also the effect of the treatment on a particular quantile q of the distribution:

$$\tau_{q^{CIC}} = F_{Y^I,11}^{-1}(q) - F_{Y^N,11}^{-1}(q) = F_{Y^I,11}^{-1}(q) - F_{Y,01}^{-1}(F_{Y,00}(F_{Y,10}^{-1}(q)))$$

The setting can be extended to the case of multiple control groups, as in my case, so that the counterfactual distribution of the treated individuals in the post-treatment period can be recovered using both groups, or just one of the two.

In my case there are 6 groups: treated classes (pre and post); control classes in PQM schools (pre and post); control classes in control schools (pre and post). To account for the fact that the 6 groups may have different distribution of covariates, I developed a weighting scheme, based on the paper of Abadie (2005), such that the distribution of all observables characteristics, both at the school and at the student level, is the same across the different groups. Defining $C = 1$ as treated classes, $S = 1$ as PQM schools, and $T = 1$ as post-treatment year, for each unit i I recover a weight, w_{CST}, such that[3] :

$$w_{CST} f_{X|CST}(c, s, t) = f_{X|CST}(x|1, 1, 1),$$

where $f_{X|CST}(x|c, s, t)$ is the distribution of covariates in group $C = c, S = s, T = t$. Therefore in the implementation of the CIC I use the weighted cumulative distribution functions.

4.3 Empirical results

4.3.1 Threshold difference-in-differences

I start presenting the results form the threshold difference-in-differences. The effect on the probability of answering correctly to at least a fixed number

[3]Details of the calculation of weights are in the Appendix

of questions and of scoring above a certain threshold at the test score are presented in Figure 4.1 and 4.2, respectively. These effects are calculated using thresholds that represent percentiles of the outcome distribution in the pre-programme year.[4] Thus, we report the treatment effect on the probability of being at least as good as the τ^{th} percentile of the outcome distribution in the pre-programme period.

Table 4.1: Correspondence between percentiles and percentage of correct answers and test scores

	Mathematics		Language	
Percentile	Percentage of correct answers	Test score	Percentage of correct answers	Test score Test score
1	0.265	-1.083	0.371	-1.234
5	0.324	-0.837	0.446	-0.835
10	0.355	-0.702	0.476	-0.653
20	0.397	-0.498	0.511	-0.457
30	0.427	-0.356	0.539	-0.290
40	0.448	-0.238	0.560	-0.151
50	0.474	-0.107	0.583	-0.020
60	0.498	0.040	0.603	0.118
70	0.529	0.217	0.625	0.269
80	0.567	0.434	0.652	0.449
90	0.624	0.789	0.683	0.691
95	0.677	1.089	0.711	0.863
99	0.760	1.698	0.773	1.341

Presented in the table are values of percentage of corrected answers and test scores associated to the percentiles. Percentiles calculated using data at the class level.

Results show that, in the classes belonging to schools in the more disadvantage background, the intervention significantly raises the probability that a class answers correctly to more than the value corresponding to the 40^{th} percentile of the baseline distribution, which correspond to 45% of correct answers in mathematics. The effect looses significance between the 55^{th} and the 70^{th} percentile, but from the 70^{th} to the 90^{th} it is again significant, even if decreasing. No effect of the intervention is found on the other two groups, and on the language percentage of correct answers. As for the other outcome, we see that the intervention increases the probability that a class test score is above all value of the baseline distribution from the 40^{th} percentile onward (in the first group of schools). Same conclusion as before hold for the other two groups and for the language test score.

[4]Values of the percentage of correct answers and test score associated to each percentile are presented in Table 4.1 for distribution at the class level and in Table 4.2 for distribution at the student level. Notice that the percentiles are calculated using the whole distribution of outcomes in pre-intervention, thus including all the students and classes in the three groups of schools.

Table 4.2: Correspondence between percentiles and percentage of correct answers and test scores

	Mathematics		Language	
Percentile	Percentage of correct answers	Test score	Percentage of correct answers	Test score Test score
1	0.143	-1.675	0.207	-2.172
5	0.214	-1.355	0.310	-1.571
10	0.262	-1.150	0.379	-1.237
20	0.333	-0.871	0.448	-0.824
30	0.381	-0.629	0.517	-0.513
40	0.429	-0.386	0.552	-0.243
50	0.476	-0.145	0.603	0.016
60	0.524	0.137	0.638	0.282
70	0.595	0.468	0.672	0.573
80	0.643	0.867	0.724	0.900
90	0.738	1.420	0.776	1.339
95	0.810	1.852	0.810	1.669
99	0.905	2.524	0.879	2.258

Presented in the table are values of percentage of corrected answers and test scores associated to the percentiles. Percentiles calculated using data at the student level.

These results suggest that not all the classes involved are actually significantly increasing their average outcomes: classes that on average were performing at the very bottom of the distribution of test scores and of percentage of correct answers, seem not to benefit from the intervention since we do not see an increase of the proportion of classes scoring above the first $(10, 20, 30^{th})$ percentiles. Thus we can say that classes that were scoring below the 40^{th} percentile of the test scores distribution, do not manage to increase their performances. These classes are probably characterised by an high proportion of low performing students. In order to have an insight of how the intervention affects single student achievement I also estimate the effect probability that each student's score (percentage of correct answers) is above a given value.

Results for schools in the bottom tertile, show that the extra time dedicated to mathematics activities increases the probability that students' percentage of correct answers (Figure 4.3) and test score (Figure 4.4) is above the value corresponding to the 30^{th} percentile of the baseline distribution of percentage of correct answers (which correspond to 38 % of correct answers), and above the value corresponding to the 40^{th} percentile of the baseline distribution of test score. In both cases the effect is stable up to the 70^{th} percentile and then it decreases. Thus the average positive effect found with the standard difference in difference is mainly driven by an increase of the probability of scoring values of y in the middle part of the distribution of scores. Nevertheless no effect is found at the very beginning of the distribution, i.e. lower performing students, do not benefit at all, since we see an higher proportion of students scoring at least the 40^{th}, but not an higher proportion of students scoring above the $10, 20, 30^{th}$ percentile.

As usual we do not find any effect in the other two groups of schools for mathematics and any effect of language activities on probabilities of scoring above given values of the language's outcomes.

Since we found that that extra hours in language had a negative average effect on test scores in mathematics, in schools belonging to the top tertile, we investigate whether the effect is constant for all the students in the class or whether it is driven just by some of the students. Thus in Figure 4.5 we plot the coefficients associated to receiving extra time in language in the mathematics equation, i.e. the effect of PQM language on the mathematics outcomes. From the figure, referring just to the schools in the top tertile, we see that the negative effect is driven by children in the top part of the outcome distributions, meaning that less students score at least as high as the 70^{th} percentile.

The estimates calculate so far refer to the intervention effect at a given level of y, and have to be interpreted as whether the intervention managed to increase the proportion of classes or students with test scores or percentage of correct answers above a given value of y. Nonetheless these estimates are not informative of which part of the distribution is benefiting the most. In order to find out differences in returns at different points of the outcome distribution we need to calculate quantile treatment effects, which are informative of the gain that a student (class) at a given quantile of the pre-intervention distribution would receive from the intervention (assuming rank invariance).

Therefore I calculate the quantile treatment effects at the student level, simply inverting the estimates from the threshold difference-in-differences estimated for the test scores (thus not considering percentage of correct answers). In figure 4.6 I plot the quantile treatment effects in the three groups of schools for mathematics and language and the corresponding 95% confidence intervals. Assuming rank invariance, we see that students who were scoring above the 40^{th} quantile of the baseline distribution receive a positive effect from the intervention in mathematics in the first group of schools. In addition the quantile effect is larger at the higher quantiles of the distribution: top performing students are the ones who receive the larger returns from the program, while very low achieving students are not affected at all.

In addition I repeated the same analysis considering as outcome variables test scores in the mathematical *reasoning* and *knowledge*. Thus I estimates the probability that a student test score in either outcomes is above a given value of y, and the corresponding quantile treatment effects. As for the mathematical *reasoning* we see that students involved in the intervention in the bottom tertile group of schools have an higher probability of scoring all values of the baseline distribution above the 30^{th} percentile, and that the effect is increasing up to the 80^{th} percentile and then it decreases. The corresponding quantile treatment effects are positive and significant after the 30^{th} and increase, with higher effect at higher quantiles.

Interestingly we notice that also in the *knowing* part of the test, enrolled students show an higher probability of scoring between the 50^{th} and 70^{th} percentile of the baseline distribution and around the 80^{th} (in the bottom tertile group of schools). Results that hold for the quantile effects, which are significant around the same quantiles. Thus, even if I was not finding any effect of the intervention on average performances in mathematical *knowledge* (See Table 3.13), there are some students who are actually improving also in that part of the test. Probably their gain is not enough to significantly increase the average class performances.

Finally, I also estimate the effect of treatment intensity on the probability that a student's test score (leaving aside percentage of correct answers) is above a given value of y. I thus replicate the analysis done in equation 3.2, but using as outcome variable a dummy which takes value 1 if student's test score y_{ijt} is above a given value of the baseline distribution. In Figure 4.9 I plot the estimated coefficients evaluated at mean values of the two measures of intensity in the three groups of schools. The figure shows that increasing the average class instruction time in mathematics by 30% increases the probability that a student scores above each value of the baseline distribution from the 40^{th} percentile onward. The same result hold for increasing the instruction time per student by 17%.

4.3.2 Change-in-changes

As a further check I provide also results from the change-in-changes model. In these estimates I consider only student level test score distribution in mathematics, thus not considering language.

As explained in Paragraph 4.2 the change-in-change model allows to recover the full counterfactual distribution of the treated group in absence of the treatment. Given the setting of the model I will first consider as control group only non treated classes in PQM schools, and then only control classes in non PQM schools. As mentioned before the distributions of the test scores have been weighted so to balance the difference in covariates between the different groups.

In Figure 4.10 I report the counterfactual and the factual distributions of the treated group in the post-intervention year, in the three groups of schools.

In the left panel I used as controls non treated classes in PQM schools, while in the right panel I used classes in non PQM schools. Clearly the two control groups are comparable, since the counterfactual distributions obtained using the two different controls are very similar.[5]

As mentioned in Paragraph 4.2, once obtained the counterfactual distribution of the treated group in absence of the treatment, it is very easy to recover both average treatment effect and quantile treatment effects.

Given the previous findings, I concentrate on the group of schools belonging to the bottom tertile. I calculate the change-in-changes estimates of receiving extra time in mathematics on mathematics test score using as control group firstly only the non treated classes in PQM schools, and secondly only the classes in non PQM schools. The average treatment effect obtained with the CIC using the first control group is 0.291 (se = 0.045) and using the second group is 0.300 (se = 0.042)[6]. We notice that the two estimates are very similar between each other, meaning that the effect of being a treated class is the same using as control either one of the two control groups and in addition we notice that these estimates are very similar to the estimate found using the standard difference-in-differences model, which was 0.257 (se = 0.111), as reported in Table 3.12. They are not exactly the same due to the different assumptions involved in the two models, and due to the fact that with the CIC I'm not able to take into account the fact that some classes used as control may be receiving extra instruction time in language and that may have an effect on the resulting estimates.

Figure 4.11 presents the quantile treatment effects : the left panel present results calculated using as control group non treated classes in PQM schools, while the right panel presents results calculated using as control group classes in non PQM schools. The solid lines represent the quantile treatment effect, while the dashed lines the bootstrap 95% confidence interval for each estimate. The quantile treatment effects estimated with the two different control groups are very similar between each other. Moreover the results found are in line with the estimates from the threshold difference-in-differences: we see that nothing is found for very low achieving students, since the effect for the very low quantiles is 0; the positive effect starts to emerge from the 40^{th} onward, and it increases over the rest of the distribution, with high performing students receiving the grater benefits from the intervention.

[5]This fact is in line with the results I have always found about not finding evidence of spill over effects on non treated classes in PQM schools.

[6]All the standard errors are bootstrapped.

4.4 Discussion and Conclusion

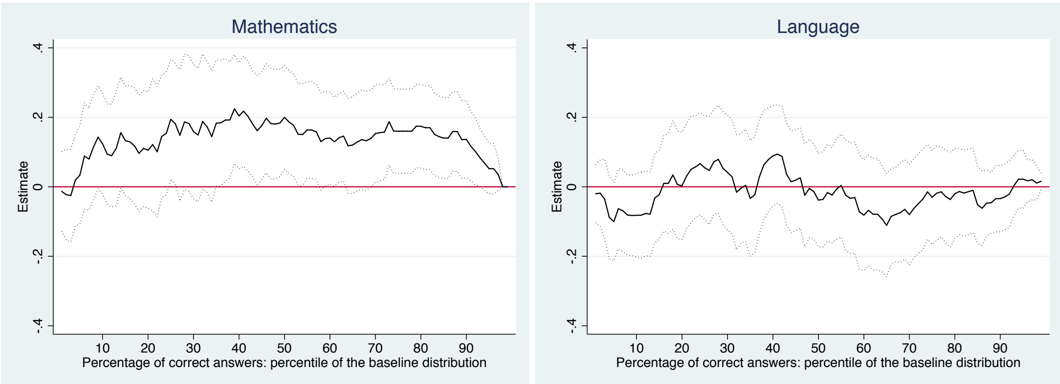
In this chapter I provide evidence about the non-linear effects of the PQM program, indeed, while average treatment effects are enough to conclude whether the intervention on average increases or decreases the performances of the treated classes, they can not answers questions about the heterogeneity of returns across the outcome distribution, i.e. they do not tell whether all the classes and students are actually benefiting from the program and whether everybody is benefiting by the same amount.

In order to answers these questions I rely on two different methods, the threshold difference-in-differences and the change-in-changes. Both methods allow to recover the effects of the intervention on the different quantiles of the test score distribution. Even if the two methods rely on slightly different assumptions, they lead to the same conclusions: in the group of schools characterised by a lower socio-economic background, the average positive effect found is driven by larger returns for children laying in the upper part of the outcome distribution, that is high performing students. On the other side, for low achieving students, we do not find any significant effect of the intervention. Thus only the students who were performing well enough (above the 40^{th}) benefits from receiving extra instruction time in mathematics, and the effect increases with the quantiles, which means that the effect is larger the "better" is the student.

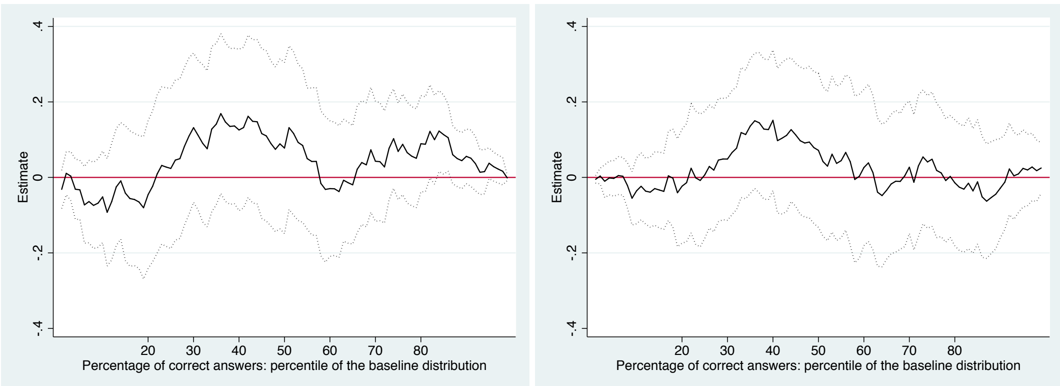
This finding may be linked to the fact that, as mentioned in the previous chapter and as is confirmed also in this chapter, the greater returns were coming from increased performances in the *reasoning* part of the mathematics test. Therefore we may expect that students with higher ability (i.e. higher performing students) may be the ones who can better exploit the extra time at school dedicating time to apply their knowledge and boosting their thinking abilities. On the other side, for the low ability students, extra classes hold in the afternoon may not be enough to stimulate and actually improve their mathematical thinking. Probably the nature of the intervention, which foresees extra activities to be held to small groups, it is not the best solution for very low achieving students: maybe for them the best would be one-to-one extra instruction time, focused on their personal needs and problems.

Figure 4.1: Probability that a class average percentage of correct answers is above y

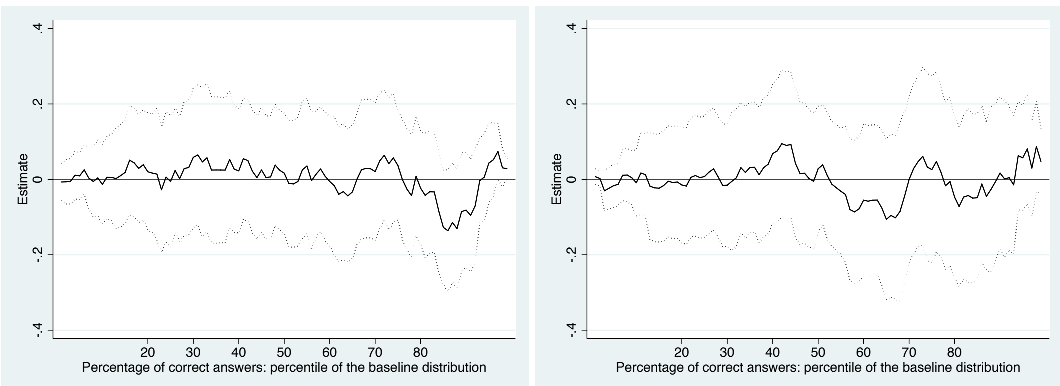
Bottom tertile



Middle tertile



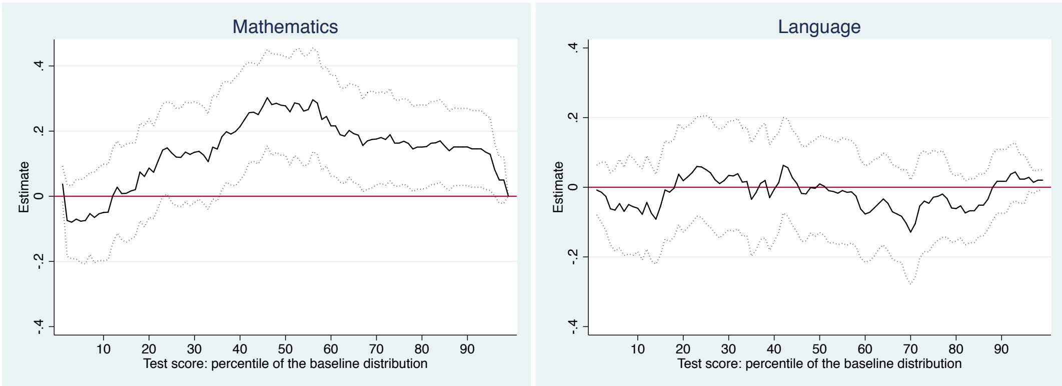
Upper tertile



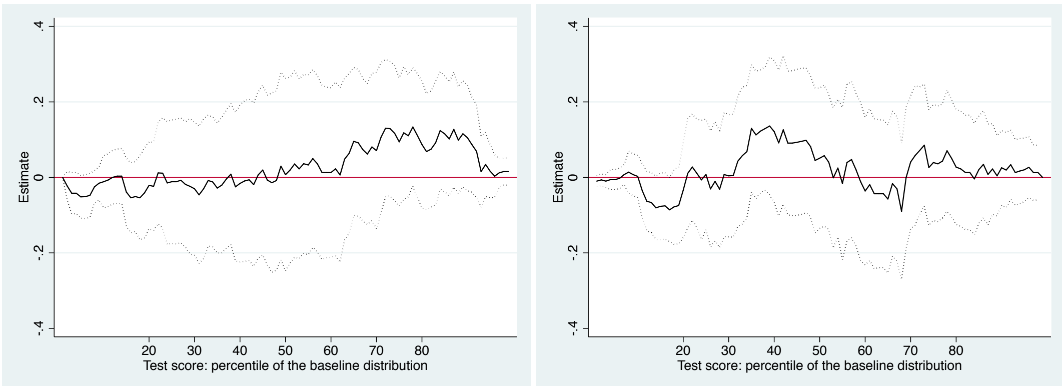
Threshold difference-in-difference estimates of the probability that a class's percentage of correct answers in mathematics (left panels) or language (right panels) is above y.

Figure 4.2: Probability that a class average test score is above y.

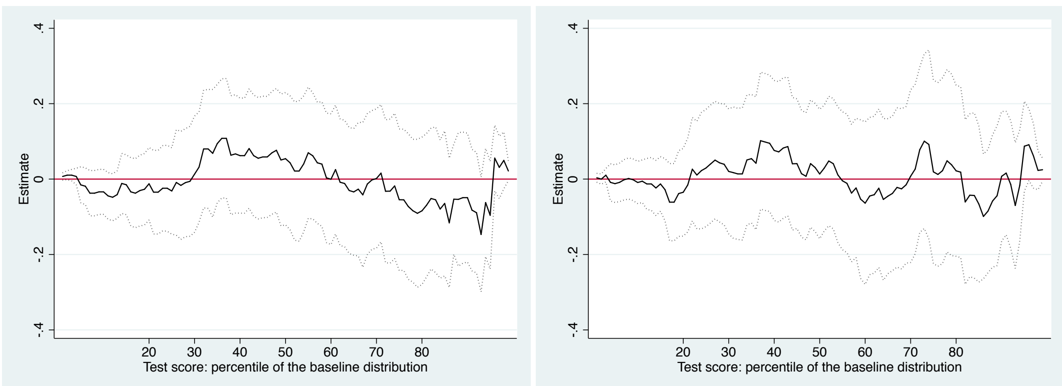
Bottom tertile



Middle tertile



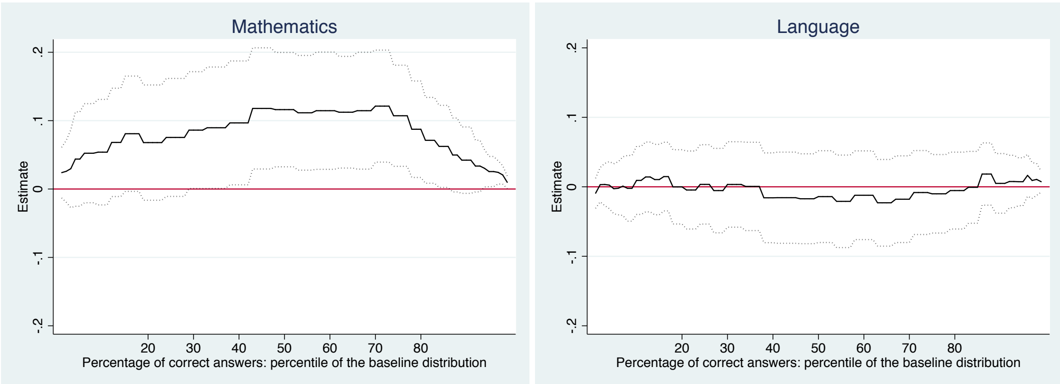
Upper tertile



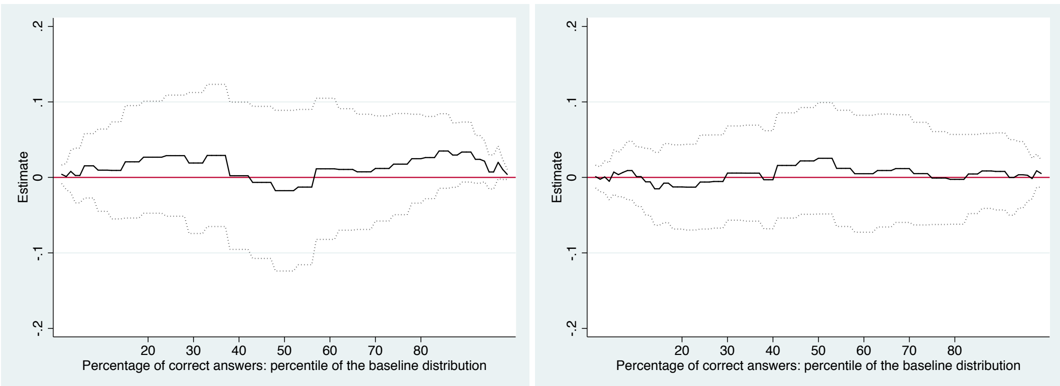
Threshold difference-in-difference estimates of the probability that a class's test score in mathematics (left panels) or language (right panels) is above y.

Figure 4.3: Probability that a student's percentage of correct answers is above y.

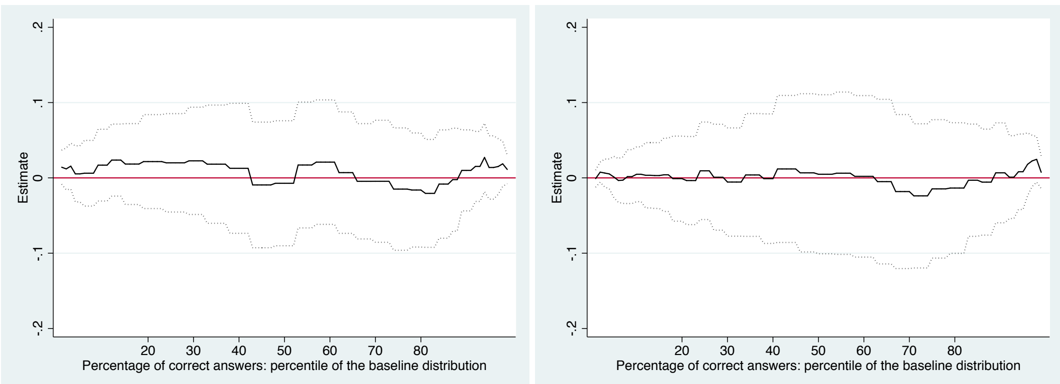
Bottom tertile



Middle tertile



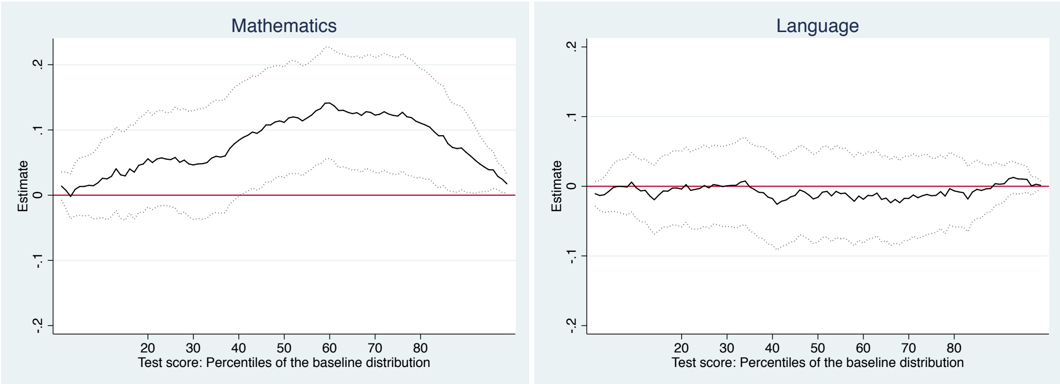
Upper tertile



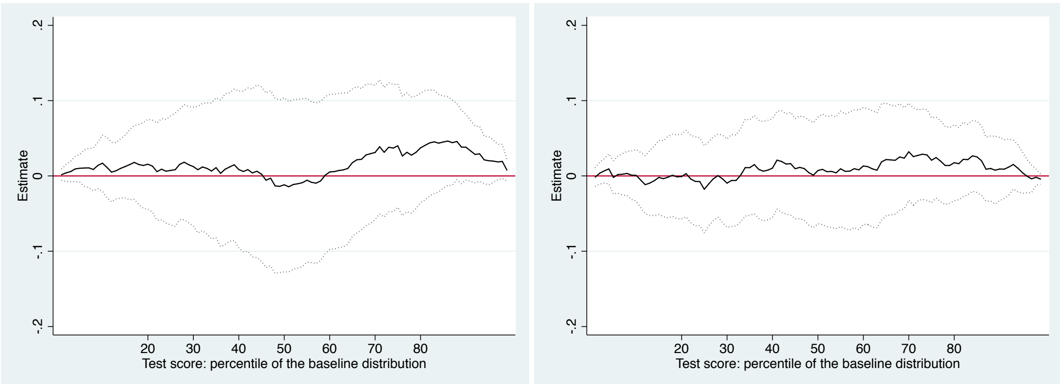
Threshold difference-in-difference estimates of the probability that a student's percentage of correct answers in mathematics (left panels) or language (right panels) is above y.

Figure 4.4: Probability that a student's test score above y.

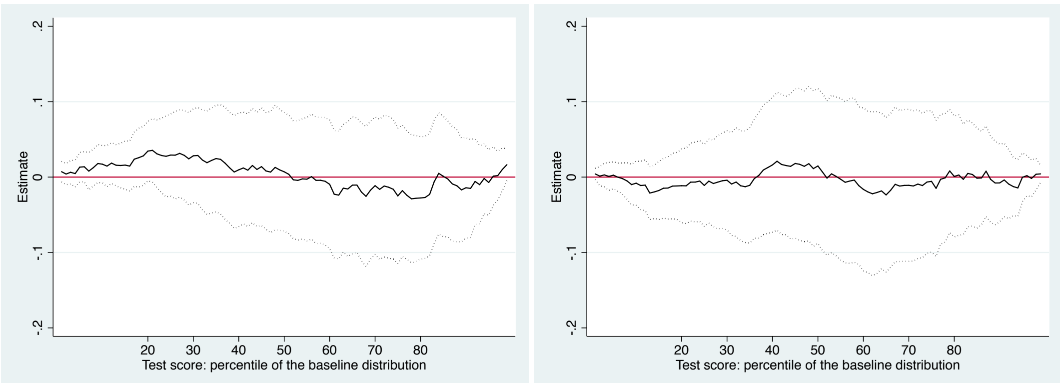
Bottom tertile



Middle tertile

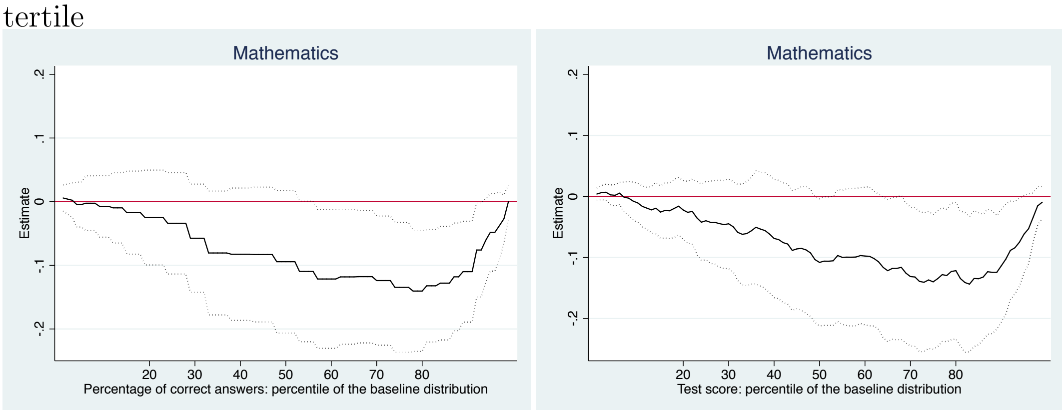


Upper tertile



Threshold difference-in-difference estimates of the probability that a student's test score in mathematics (left panels) or language (right panels) is above y.

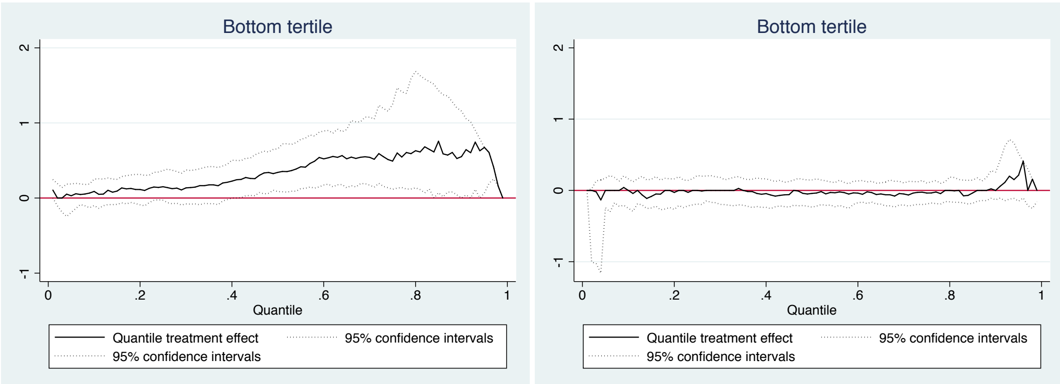
Figure 4.5: Probability that a student outcome is above y: effect of extra time in language on mathematics in the group of schools belonging to the upper tertile



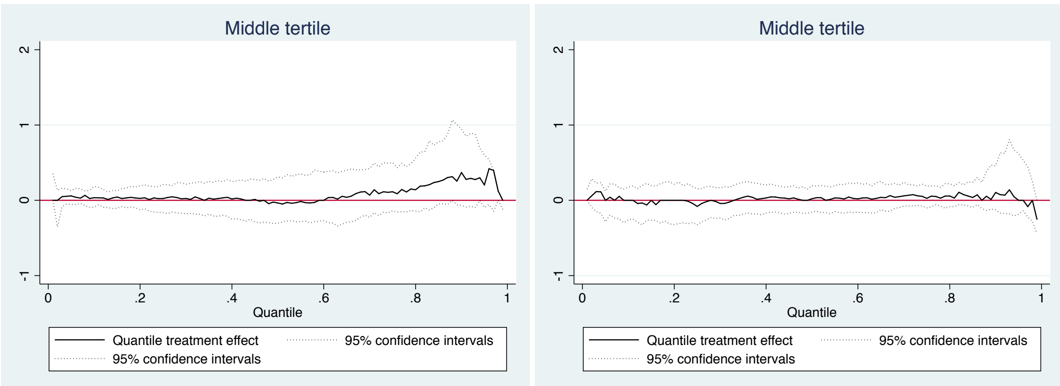
Threshold difference-in-difference estimates of being enrolled in any extra activity in language on the probability that a student's mathematics percentage of correct answers (left panel) and test score (right panel) is above y. Figure refers to the school in the top tertile of pre-treatment year test score distribution.

Figure 4.6: Quantile treatment effect on mathematics and Italian language test score

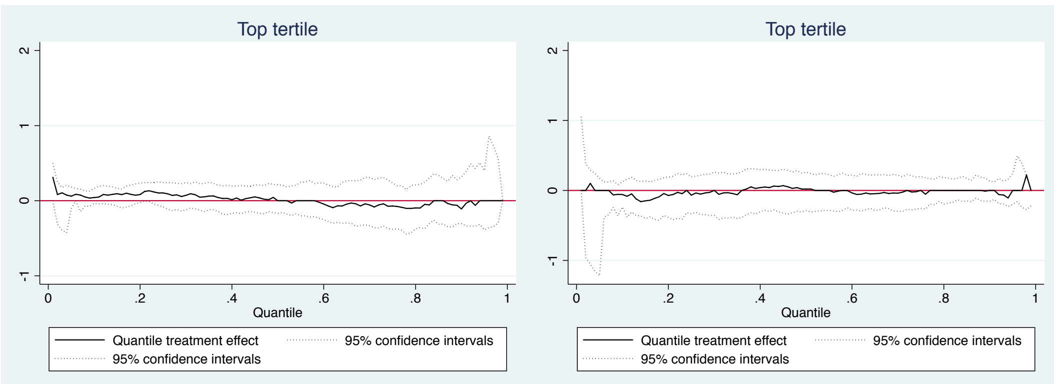
Bottom tertile



Middle tertile



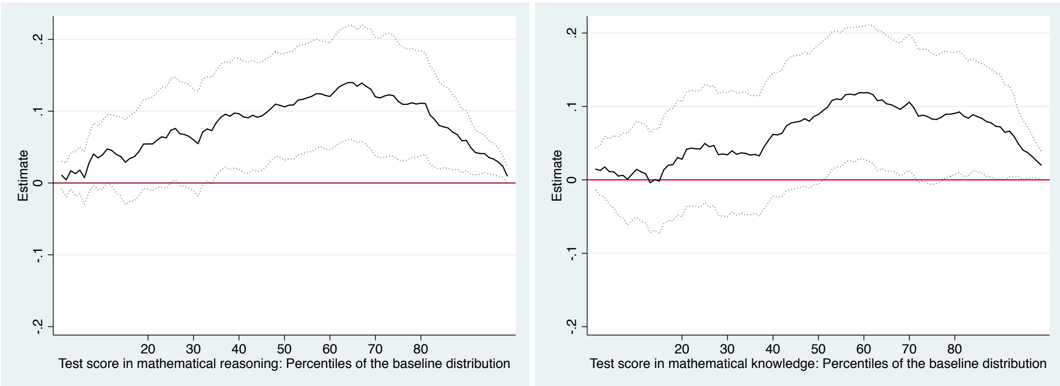
Top tertile



Quantile treatment effects for the mathematics (left panels) and language (right panel) test scores in the three groups of schools. Quantile effects have been calculated by inverting the estimates of the threshold difference-in-difference estimates of the probability that a student's test score is above y.

Figure 4.7: Probability that a student's test score above y in the mathematical reasoning and mathematical knowledge.

Bottom tertile



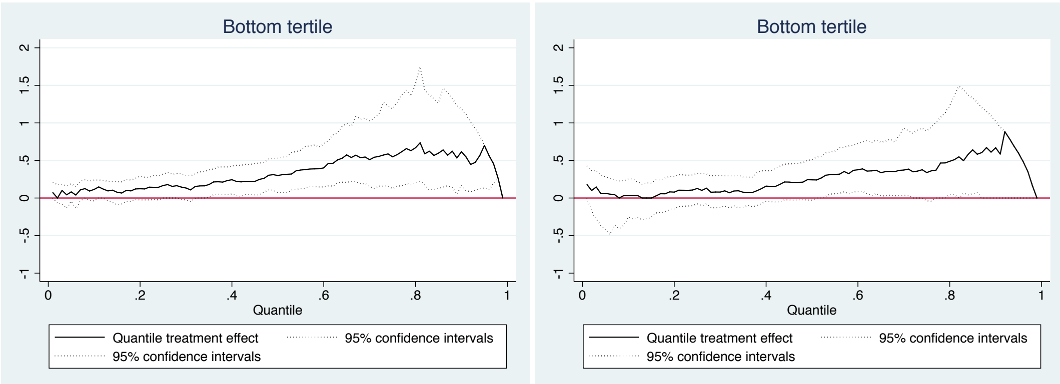
Middle tertile

Upper tertile

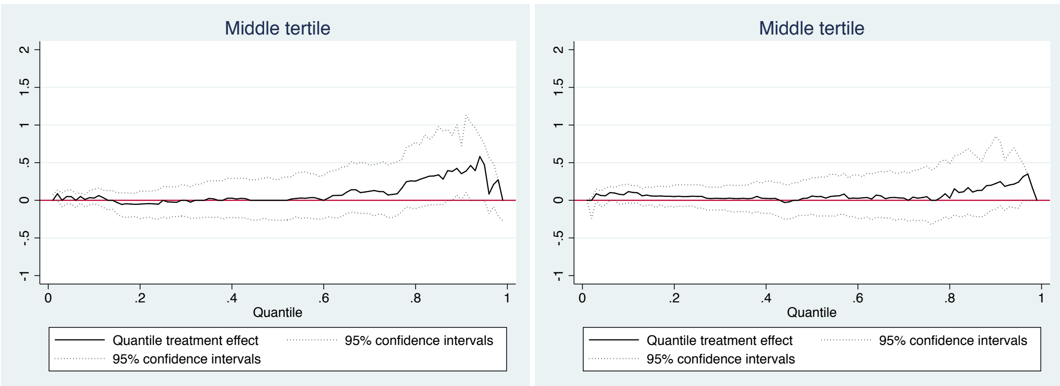
Threshold difference-in-difference estimates of the probability that a student's mathematics test score is above y, in the mathematical reasoning (left panels) and mathematical knowledge (right panels), in the three groups of schools.

Figure 4.8: Quantile treatment effects in mathematical reasoning and mathematical knowldege.

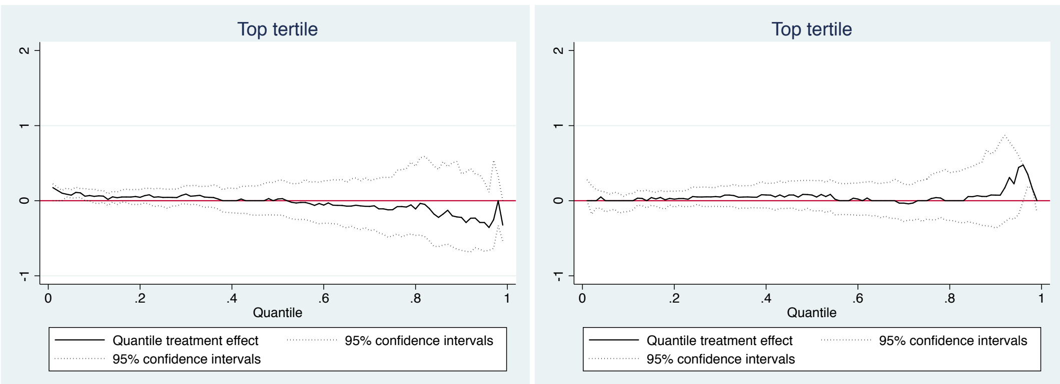
Bottom tertile



Middle tertile



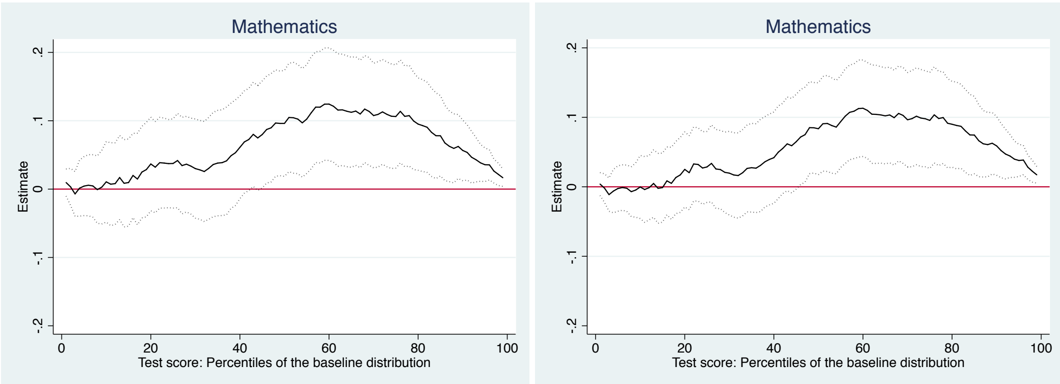
Upper tertile



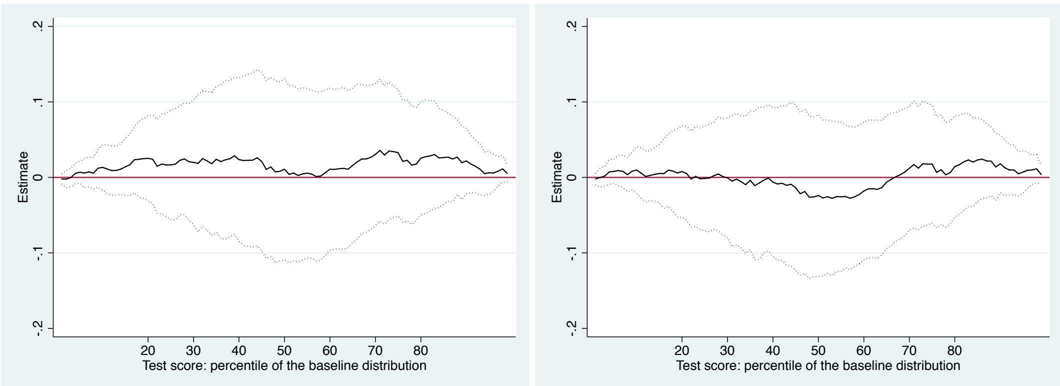
Quantile treatment effects for the mathematical reasoning (left panel) and mathematical knowledge (right panels) in the three groups of schools. Quantile effects have been calculated by inverting the estimates of the threshold difference-in-difference estimates of the probability that a student's mathematics test score in the two parts of the test is above y.

Figure 4.9: Probability that a student's test score above y: effect of treatment intensity on the test score in mathematics.

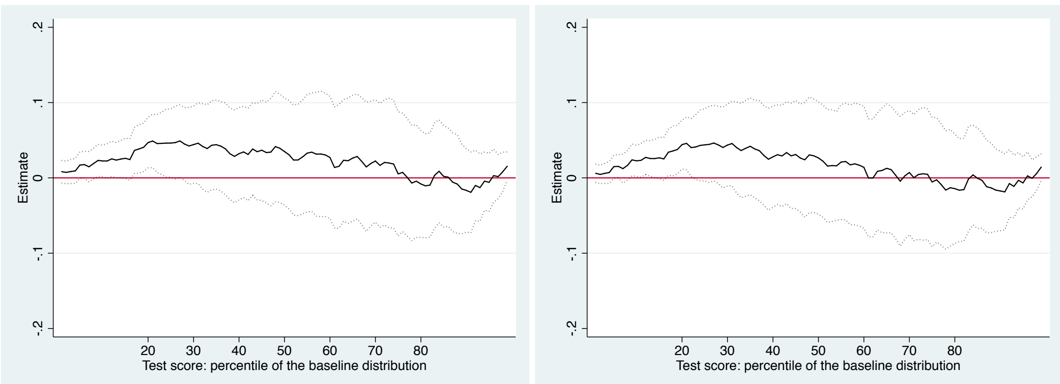
Bottom tertile



Middle tertile



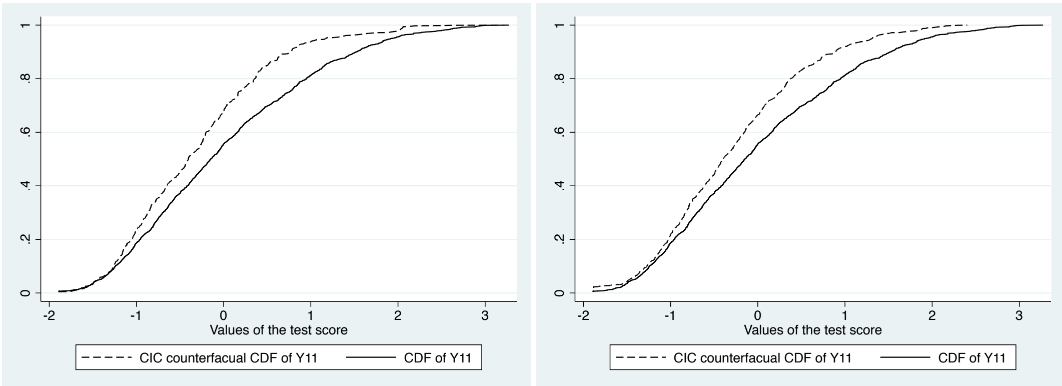
Upper tertile



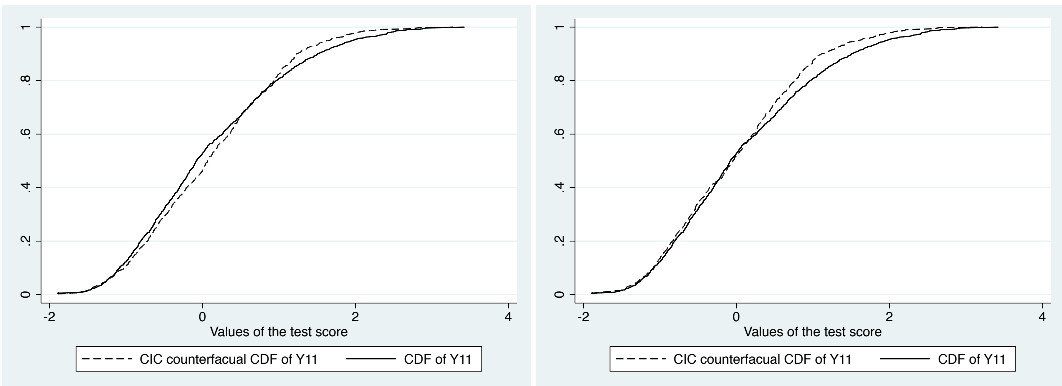
Threshold difference-in-difference estimates of the probability that a student's mathematics test score is above y, using as main dependent variables treatment intensity. In the left panels, intensity is defined as the percentage change in instruction time, in the right panels as the percentage change in instruction time per student. The estimates correspond to mean value of intensity in the different groups.

Figure 4.10: Counterfactual distributions of the treated group

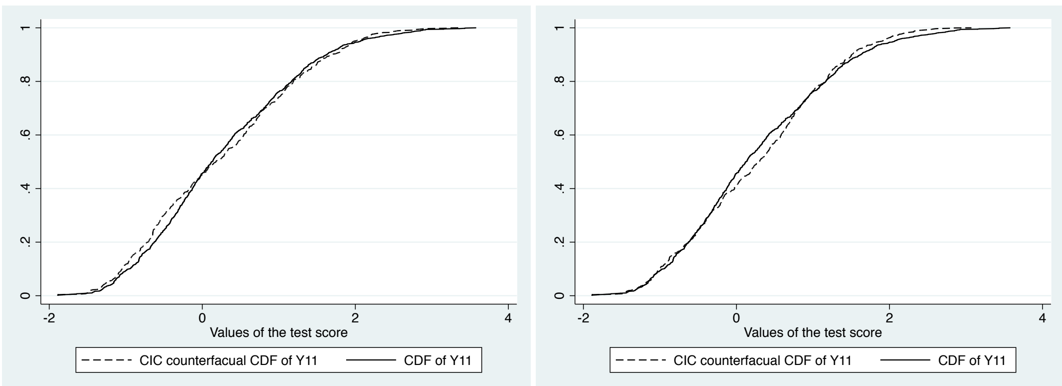
Bottom tertile



Middle tertile

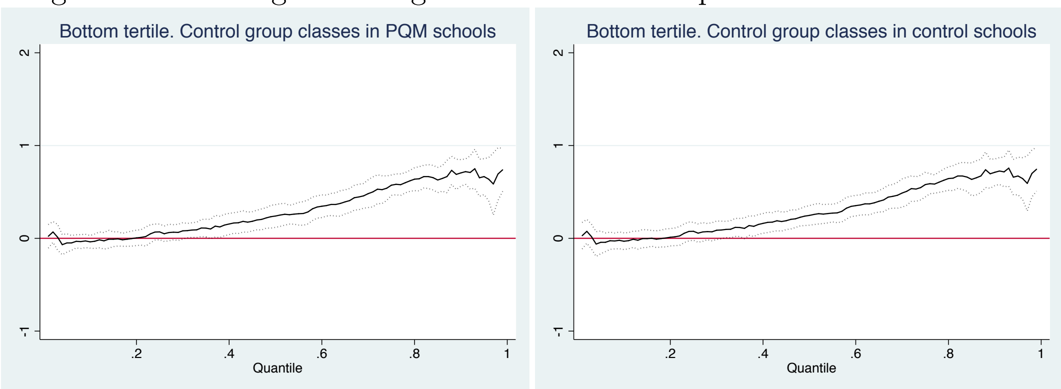


Upper tertile



Counterfactual and factual distributions of the test score in mathematics in the three groups of schools. Counterfactual distributions are calculated using the Change-in-changes method. The left panels use as control group non treated classes in PQM schools, while the right panels use classes in non PQM schools .

Figure 4.11: Change-in-changes estimates of the quantile treatment effects



The graph presents the changes-in-changes estimates of the impact on the mathematics test score in the group of school belonging to the bottom tertile for each quantile of the distribution. The left panel uses as control group non treated classes in PQM schools, while the right panel uses classes in non PQM schools . The dashed lines represent the bootstrapped 95% confidence interval for each estimate.

Conclusions

The aim of this thesis is to provide evidence on the effectiveness of a programme providing students in low achieving schools with extra instruction time in mathematics and Italian language. The intervention has been financed by the European Social Fund, and it targets low achieving lower secondary schools in Southern Italy. Schools in this areas are characterised by low performances in both mathematics and reading, compared to schools located in the Northern part of the country, in addition international surveys have highlighted that student performances in this area are much below the OECD average. Thus it is important to shed light on the effectiveness of interventions such the one proposed, and verify whether children living in these environments could be helped to catch up with their European and Italian colleagues. Moreover this thesis is filling the current gap about the effectiveness of EU investments made in the most deprived areas of the Italian country.

The Quality and Merit programme (PQM) increases the number of hours children spend at school, in particular increasing either the mathematics and/ or Italian language hours in a selected number of classes of the schools chosen to participate. With a matching, I select a group of control schools similar with respect to a long list of pre-programme indicators to the enrolled ones, and using repeated observation of the same *sezione* over time, I am able to take into account sorting of children and teachers in the classes. Indeed to control for sorting across classes I use the fact that student are divided into groups distinguished by letters (*sezione*), that they remain in the same group across grades and that the composition of teachers in the school assigned to each group is substantially stable over time. I implement a difference-in-differences strategy, and compare two contiguous cohorts of sixth grade students enrolled in the same group. I contrast groups with and without additional instruction time in participating schools, to groups in non-participating schools that I selected.

In Chapter 3 results show that extra time at school spent doing mathematics activities increases the test score and the percentage of correct answers in mathematics, just in schools characterised by lower pre-intervention performances (schools whose average performances in the pre-treatment year where below the first tertile of the distribution of test scores), while no effect is found for treated classes in schools belonging to the other two tertiles. I believe that children belonging to the first tertile group are the ones coming from the lower socio-economic background, and I interpret the positive effect found in mathematics in two ways: the extra time they spend during the PQM programme is the only time they dedicate to study outside regular school time, thus they are actually spending more time on academic activities, which means that achievement works as a cumulative process, and more time at schools results in higher performances. In addition spending more time at school also decreases the "negative" influence of the families, which I assume are not very supportive and helpful for the children in context characterised by low socio-economic background. Thus, children involved in the intervention spend at school the time they would otherwise spend doing nothing at home. On the other side, I think that in environments with higher socio-economic background, the PQM programme is working as a substitute of the work the children would anyway do at home, thus it is not effective, since also without the programme, the children would probably dedicate some time in their afternoons to academic activities.

The effect is mainly driven by returns in the part of the test measuring ability to use mathematical reasoning, thus the ability to use the knowledge of mathematical concepts and to apply them to solve problems, meaning that extra instruction time does not increases the pure knowledge of mathematics concepts, but it can help the students to improve their ability to apply their knowledge and to develop the cognitive processes needed in the mathematical reasoning.

I also investigate whether different exposure to the treatment had any different effect on mathematics outcome, thus I define two measures of treatment intensity, exploiting variability in the number of activities done in the afternoon and in the number of students involved in each activity. Using both measures I find a positive effect on mathematics test score in the first group of schools, thus the more the children are exposed, the more they seem to learn.

In addition it is interesting that the effect is found just for mathematics, while nothing it is found for Italian language. Nevertheless this results it is not new in the literature, since it has been shown that is much harder to intervene on reading and comprehension skills, rather than on skills involving a lot of exercise and practice, such as mathematics. On the other side, I find that extra time spent at school doing Italian language activities has a negative effect on mathematics test score in the top tertile group of schools. This result suggests that additional time spent at school in reading activities may substitute the time the children would have otherwise invested in studying mathematics.

In Chapter 4 I investigate non-linear effects, going beyond averages, thus I rely on two methods, that allow me to understand whether all the classes and all the student are benefiting in the same way, or whether there are classes or students with higher or lower returns. Using the first method, the "threshold difference-in-difference", I estimate the probability that a class average test score is above a given value of y. Results show that the effect is positive and significant from the 40^{th} percentile onward: treated classes have an higher probability of scoring at least a value of y corresponding to the 40^{th} percentile onward. This means that not all the classes are actually benefiting from this intervention, but, if we assume rank invariance, the classes benefiting are the ones with test score above the 40^{th} percentile. When I look at the probability that a student's test score is above a given value of y I find that the intervention increases the probability that test scores are above the 40^{th} percentile of the baseline distribution, and that the effect is increasing up to the 70^{th} percentile and then it decreases. Calculating the corresponding quantile treatment effects, I notice that the effect is larger for higher quantiles, thus the students who receive larger returns are the ones at the very top of the test score distribution, while no effect is found for very low achieving students. These results are confirmed also using the second methods, the "change-in-changes".

In terms of policy implication I could suggest that intervention such the one evaluated can be effective in very deprived environments, where children have no other options after schools. Here any more time spent at school could be helpful since is the only way to make pupils study more. Nevertheless this intervention was not successful for all: students in the very bottom part of the distribution, i.e. worse students, do not seem to benefits. Probably for these students extra instruction time it is not enough, and they would need a really targeted intervention focusing just on them, probably in smaller groups and for longer hours.

Finally the intervention successfully raised only mathematics test scores, while no gains are fund for Italian language test scores. This could suggest that extra instruction time for children in the considered age (11-12 years old) is effective only when providing help in subjects which involve exercises and practice, like mathematics or sciences, while I think it is not enough to increase reading skills. Probably, as documented in previous studies, in order to actually affect reading skills we should target younger children, during elementary schools.

Appendix

Calculation of the weights used in the CIC model

Observations are divided in 6 different groups. In what follow let C^k be a dummy for treated classes in subject k in PQM schools, S be a dummy for PQM schools, and T a dummy for the post-program period.

For the two subjects $k =$ (mathematics, language) I calculate weights w_{CST} so that the distribution of all observable characteristics is the same across the 6 groups.
Let $X = [X_1, X_2]$ be the set of variables used. The set X_1 refers to variables at the school level, while X_2 refers to variables at student level.
I will weight observations to match the distribution of X in the $(C = 1, S = 1, T = 1)$group. So that:

$$w_{CST} f_{X|CST}(x|c,s,t) = f_{X|CST}(x|1,1,1)$$

The weights in the 5 groups are:
$(C = 1, S = 1, T = 0)$

$$\begin{aligned}
w_{110} &= \frac{f_{X|CST}(x|1,1,1)}{f_{X|CST}(x|1,1,0)} \\
&= \frac{f_{CST|X}(1,1,1|x)}{f_{CST|X}(1,1,0|x)} \frac{f_{CST}(1,1,0)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(0|1,1,x)} \frac{f_{CST}(1,1,0)}{f_{CST}(1,1,1)}
\end{aligned}$$

$(C = 0, S = 1, T = 1)$

$$\begin{aligned}
w_{011} &= \frac{f_{X|CST}(x|1,1,1)}{f_{X|CST}(x|0,1,1)} \\
&= \frac{f_{CST|X}(1,1,1|x)}{f_{CST|X}(0,1,1|x)} \frac{f_{CST}(0,1,1)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(1|0,1,x)} \frac{f_{C|SX}(1|1,x)}{f_{C|SX}(0|1,x)} \frac{f_{CST}(0,1,1)}{f_{CST}(1,1,1)}
\end{aligned}$$

$(C = 0, S = 1, T = 0)$

$$
\begin{aligned}
w_{011} &= \frac{f_{X|CST}(x|1,1,1)}{f_{X|CST}(x|0,1,0)} \\
&= \frac{f_{CST|X}(1,1,1|x)}{f_{CST|X}(0,1,0|x)} \frac{f_{CST}(0,1,0)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(0|0,1,x)} \frac{f_{C|SX}(1|1,x)}{f_{C|SX}(0|1,x)} \frac{f_{CST}(0,1,1)}{f_{CST}(1,1,1)}
\end{aligned}
$$

$(C = 0, S = 0, T = 1)$

$$
\begin{aligned}
w_{011} &= \frac{f_{X|CST}(x|1,1,1)}{f_{X|CST}(x|0,0,1)} \\
&= \frac{f_{CST|X}(1,1,1|x)}{f_{CST|X}(0,0,1|x)} \frac{f_{CST}(0,0,1)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(1|0,0,x)} \frac{f_{C|SX}(1|1,x)}{f_{C|SX}(0|0,x)} \frac{f_{CST}(0,0,1)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(1|0,0,x)} f_{C|SX}(1|1,x) \frac{f_{S|X}(1|X)}{f_{S|X}(0|X)} \frac{f_{CST}(0,1,1)}{f_{CST}(0,0,1)}
\end{aligned}
$$

$(C = 0, S = 0, T = 0)$

$$
\begin{aligned}
w_{011} &= \frac{f_{X|CST}(x|1,1,1)}{f_{X|CST}(x|0,0,0)} \\
&= \frac{f_{CST|X}(1,1,1|x)}{f_{CST|X}(0,0,0|x)} \frac{f_{CST}(0,0,0)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(0|0,0,x)} \frac{f_{C|SX}(1|1,x)}{f_{C|SX}(0|0,x)} \frac{f_{CST}(0,0,0)}{f_{CST}(1,1,1)} \\
&= \frac{f_{T|CSX}(1|1,1,x)}{f_{T|CSX}(0|0,0,x)} f_{C|SX}(1|1,x) \frac{f_{S|X}(1|X)}{f_{S|X}(0|X)} \frac{f_{CST}(0,0,0)}{f_{CST}(1,1,1)}
\end{aligned}
$$

Bibliography

A. Abadie. Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19, 2005.

L.S. Aiken, S.G. West, D.E. Schwalm, J.L. Carroll, and S. Hsiung. Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation. *Evaluation Review*, 22(2):207–244, 1998.

J.D. Angrist and V. Lavy. Does teacher training affect pupil learning? evidence from matched comparisons in jerusalem public schools. *Journal of Labor Economics*, 19(2):343–369, 2001.

S. Athey and G.W. Imbens. Identification and inference in nonlinear difference-in-difference models. *Econometrica*, 74(2):431–497, 2006.

E. Battistin and A. Schizzerotto. Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in italy. *IZA Discussion Paper*, 2013.

C. Bellei. Does lengthening the school day increase students academic achievement? results from a natural experiment in chile. *Economics of Education Review*, 28(5):629–640, 2009.

E.P. Bettinger and B.T. Long. Addressing the needs of underprepared students in higher education. *Journal of Human Resources*, 44(3):736–771, 2009.

J.C. Calcagno and B.T. Long. The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance. *National Bureau of Economic Research*, 2008.

A. Checkoway, B. Boulay, B. Gamse, M. Caven, L. Fox, K. Kliorys, R. Luck, K. Maree, M. Velez, M. Woodford, et al. Evaluation of the expanded learning time initiative. year four integrated report: 2009-10. *Annual Report, Volume 1*, 2011.

J.S. Coleman, EQ Campbell, CJ Hobson, J. McPartland, AM Mood, FD Weinfeld, et al. Equality of educational opportunity (no. oe-38001). washington, dc: Us department of health, education, and welfare, office of education. *National Center for Education Statistics*, 1966.

P. DeCicca. Does full-day kindergarten matter? evidence from the first two years of schooling. *Economics of Education Review*, 26(1):67–82, 2007.

W. Dobbie and R.G. Fryer Jr. Getting beneath the veil of effective schools: Evidence from new york city. *National Bureau of Economic Research*, 2011.

M. Dynarski, S. James-Burdumy, M. Moore, L. Rosenberg, J. Deke, W. Mansfield, et al. When schools stay open late: The national evaluation of the 21st century community learning centers program: New findings. *Report submitted to the US Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, DC: US Government Printing Office*, 2004.

P. Falzetti, L. Palmerio, B. Romano, and D. Vidoni. Reducing the gap? the effects of eu funds for improving education in southern italy. Vancouver, British Columbia, Canada, 2012. AERA Annual meeting.

S. Firpo, N.M. Fortin, and T. Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

C.H. Gibbs. Measuring the impact of full-day kindergarten: Experimental and quasi-experimental evidence. *Society for Research on Educational Effectiveness*, 2010.

J.B. Grossman, C.L. Sipe, and Corporation for Public/Private Ventures. *Summer Training and Education Program (STEP): Report on long-term impacts*. Public/Private Ventures, 1992.

B. Hansen. School year length and student performance: Quasi-experimental evidence. *Unpublished manuscript. University of California, Santa Barbara*, 2008.

E.A. Hanushek, J.F. Kain, D.M. O'Brien, and S.G. Rivkin. The market for teacher quality. *National Bureau of Economic Research*, 2005.

T. Havnes and M. Mogstad. Is universal child care leveling the playing field?: Evidence from non-linear difference-in-differences. *IZA Discussion paper*, 2010.

J.J. Heckman, J. Smith, and N. Clements. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4):487–535, 1997.

INVALSI. Le competenze in lettura, matematica e scienze degli studenti quindicenni italiani. rapporto nazionale pisa 2009. Technical report, INVALSI, 2010a.

INVALSI. Servizio nazionale di valutazione 2010-11. le rilevazioni degli apprendimenti a.s. 2010-11. Technical report, INVALSI, 2010b.

B.A. Jacob and L. Lefgren. Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics*, 86(1): 226–244, 2004.

A.B. Krueger. Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2):497–532, 1999.

L. Lang, J. Torgesen, W. Vogel, C. Chanter, E. Lefsky, and Y. Petscher. Exploring the relative effectiveness of reading interventions for high school students. *Journal of Research on Educational Effectiveness*, 2(2):149–175, 2009.

A. Lash, S. Bae, V. Barrat, E. Burr, and T. Fong. Full-day kindergarten and student achievement: A literature review. *US Department of Education, REL West Technical Assistance Response. Washington, DC*, 2008.

V. Lavy. Do differences in schools instruction time explain international achievement gaps in math, science, and reading? evidence from developed and developing countries. *National Bureau of Economic Research*, 2010.

V. Lavy. Expanding school resources and increasing time on task: Effects of a policy experiment in israel on student academic achievement and behavior. *National Bureau of Economic Research*, 2012.

V. Lavy and A. Schlosser. Targeted remedial education for underperforming teenagers: Costs and benefits. *Journal of Labor Economics*, 23(4):839–874, 2005.

J.W. Lee and R.J. Barro. Schooling quality in a cross–section of countries. *Economica*, 68(272):465–488, 2001.

P. Mandel and B. Süssmuth. Total instructional time exposure and student achievement: An extreme bounds analysis based on german state-level variation. *CESifo Working Paper Series*, 2011.

D.E. Marcotte. Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, 26(5):629–640, 2007.

D.E. Marcotte and S.W. Hemelt. Unscheduled school closings and student performance. *Education Finance and Policy*, 3(3):316–338, 2008.

E. Meyer and C. Van Klaveren. Evaluation of an extended day program in the netherlands: A randomized field experiment. *TIER Working Papers*, 2011.

J.S. Pischke. The impact of length of the school year on student performance and earnings: evidence from the german short school year. *Economic Journal*, 117(523):1216–1242, 2007.

S.G. Rivkin, E.A. Hanushek, and J.F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

K.B. Robin, E.C. Frede, and W.S. Barnett. Is more better? the effects of full-day vs half-day preschool on early school achievement. 2006.

D.P. Sims. Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review*, 27(1):58–68, 2008.

L. Wößmann. Schooling resources, educational institutions and student performance: the international evidence. *Oxford Bulletin of Economics and Statistics*, 65(2):117–170, 2003.

R. Zimmer, L. Hamilton, and R. Christina. After-school tutoring in the context of no child left behind: Effectiveness of two programs in the pittsburgh public schools. *Economics of education Review*, 29(1):18–28, 2010.

Meroni Elena Claudia

CURRICULUM VITAE *(January 2013)*

Personal Details

Date of Birth: July 19, 1984
Place of Birth: Desio (MB), Italy
Nationality: Italian

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.
e-mail: elena.meroni@stat.unipd.it;

Current Position

Since January 2010 (expected completion: early 2013 year)
PhD Student in Statistical Sciences
Department of Statistical Sciences, University of Padova
Thesis title : "Average and quantile effects of more instruction time in low achieving schools: evidence from Southern Italy"
Supervisors: Prof. Erich Battistin

Research interests

Policy evaluation, economics of education and economics of the family

Education

September 2006-December 2008
Master of Sciences in Economic and Social Sciences
Bocconi University, Milan
Thesis title: "Postponement and Recuperation of Fertility in Germany. The Role of Education"
Supervisor: Prof. Arnstein Aassve

September 2003- July 2006
Bachelor Degree in International Economic and Management
Bocconi University, Milan

Visiting periods

February - October 2012
Visiting at Univesitat Pompeu Fabra, Department of Political and Social Sciences, Barcelona, Spain

January - December 2011
Visiting at Carlo F. Dondena Centre for Research on Social Dynamics, Bocconi University, Milan, Italy

Further Education

2011, **Multiple Equation Estimation Techniques**, PhD course, Bocconi University
Instructor: Prof. Tommaso Nannicini and Prof. Michele Pellizzari
Topics: Instrumental variables techniques; Limited dependent variable models (Probit; Logit; Tobit; Multinomial models); Panel data (Static linear models)

2010, **Topics in the Analysis of Public Policy**, Padua University
Instructor: Prof. Enrico Rettore
Topics: Experimental evaluation of a public intervention; The role of observational studies for the evaluation of a public intervention Methods based on regression, on stratification and on matching; Instrumental variables; Difference in differences; Regression-discontinuity designs

Work experience

December 2011 - June 2012
Co-investigator at the INVALSI (National institute for the evaluation of the educational system), Frascati (Italy). *"Evaluation of the PQM program on students' achievement"*

December 2008 - December 2009
Junior researcher at Carlo F. Dondena Centre for Research on Social Dynamics, Bocconi University, Milan (Italy)

Teaching activities

Tutor for "Statistics", Bachelor of Business Administration and Management, Bocconi University, academic year 2009/2010

Teaching assistant for "Quantitative Methods for Management", Master of Science in Management and Economics in Arts, Culture, Media and Entertainment, Bocconi University, academic year 2011/2012

Awards and Scholarships

January 2010 - present
Italian Ministry of Educatin, University and Scientific Research: Three-year scholarship for Ph.d. studies at the University of Padova

September 2006 - December 2008
Bocconi scholarship for outstanding students

Skills

Computer

- Microsoft Office (Word, Excel, Powerpoint) and LaTeX
- Statistical packages: advance knowledge: Stata, Aml, SPSS

Languages

- Italian mother tongue, English fluent, Spanish fluent

Refereed Journal Articles

Grandparenting and Childbearing in the Extended Family. (2012) *European Journal of Population*, 28(4): pp 499-518 (with A. Aassve and C. Pronzato)

Leaving Mum Alone? The Effect of Parental Separation on Children Leaving Home Decisions. (2012) *European Journal of Population*, 28(3): pp 337-357 (with L. Mencarini and C. Pronzato)

Working papers

Teacher training, extra education and students' achievement. The evaluation of a program for promoting students' performances in Italian lower secondary schools. (With G. Abbiati), *paper presented at the American Educational Research Association (AERA) Annual Meeting*, Vancuover, 2012

Conference Presentations

"Should we increase instruction time in low achieving schools? Evidence from souther Italy" *Fifth Italian Congress on Econometrics and Empirical Economics*, Genova, January 2013.

"Teacher Training, Extra Education and Students' Achievement: The Evaluation of a Program for Promoting Students' Performances in Italian Lower Secondary Schools." *Improving Education throughout Accountability and Evaluation. Lessons from around the World*, Rome, October 2012.

"Postponement and recuperation of fertility in Italy. The role of education." *PAA Conference*, Dallas, USA, April 2010.

"Postponement and recuperation of fertility in Italy and Germany. The role of education." *VID Conference on Education and Demography*, Vienna, Austria, December 2009.

"Grandparenting and childbearing in the extended family" *SHARE User Conference*, Mainz, Germany, October 2008.

References

Prof. Erich Battistin
Department of Statistical Sciences
University of Padova, 35121 Padova, Italy
e-mail: erich@stat.unipd.it

Prof. Arnstein Aassve
Department of Decision Sciences
Bocconi University,
Via Roentgen, 1, 20136 Milano, Italy
e-mail: arnstein.aassve@unibocconi.it