Carlo F. Dondena Centre for Research on Social Dynamics

DONDENA WORKING PAPERS

Gender differences in exposure to more instruction time. Evidence from Italy.

Elena Claudia Meroni and Giovanni Abbiati

Working Paper No. 64
July 2014
(pre-copyedit)

Carlo F. Dondena Centre for Research on Social Dynamics
Università Bocconi
Via Guglielmo Röntgen 1, 20136 Milan, Italy
http://www.dondena.unibocconi.it

ISSN-2035-2034

Gender differences in exposure to more instruction time.Evidence from Italy *

Elena Claudia Meroni †
European Commission,
Joint Research Centre (JRC),
Econometrics and Applied Statistic Unit

Giovanni Abbiati
University of Trento
Department of Sociology

Abstract

This paper investigates the short-term effects on achievement, study behaviours and attitude of an intervention providing extra instruction time to students in lower secondary schools in southern Italy. We use a difference-in-differences strategy and compare two contiguous cohorts of students enrolled in the same class for two consecutive years. We control for sorting of students and teachers across classes using the fact that, due to a recent reform, the group of teachers assigned to each class is stable over time. We find that the programme increased performances in mathematics while no effect is found for Italian language test scores, but the programme increased positive attitudes towards both subjects. We investigate the heterogeneity of the effects focusing on the gender dimension, and find that boys and girls react differently to the intervention: girls use the extra instruction time as a complement to regular home study, while boys use it as a substitute.

JEL classification: C31, J160, I28

Keywords: Education Policy, Policy Evaluation, Instruction time, Gender differences

---

*This paper greatly benefited from discussion with Erich Battistin, Daniele Vidoni, Gianluca Argentin and Aline Pennisi, and from comments by the audience, at the "Improving Education through Accountability and Evaluation" conference (Rome, October 2012) and the AERA 2012 annual meeting (Vancouver, April 2012). We also thank INVALSI (Italian National Institute for the Evaluation of Educational Systems) for providing the data and in particular Daniele Vidoni, Patrizia Falzetti, Michele Cardone and Andrea Caputo for continuous support, and INDIRE (Italian Institute for Educational Innovation, Documentation and Research), in particular Samuele Calzone, Sara Mori and Serena Greco. Elena Claudia Meroni, European Commission, Joint Research Centre, Unit DDG.01 − Econometrics and Applied Statistic, Via Fermi 2479, 21027 Ispra, Italy; mail: elena.meroni@jrc.ec.europa.eu; phone: +39 0332 785608 and Giovanni Abbiati, University of Trento, Department of Sociology, via Verdi, 26 I-38122 Trento, email:giovanni.abbiati@unitn.it. The views expressed in this article are those of the authors and should not be attributed to the European Commission.

†Corresponding author

1 Introduction

Dating back to the mid-sixties and the publication of the Coleman Report (Coleman et al., 1966), the question of how to improve skills among students has been one of the most hotly-debated topics in the social sciences. In many disciplines, most of the attention has been devoted to measuring the effect of a range of school inputs on student outcomes (typically, student achievement in core subjects such as reading, mathematics and sciences) and the heterogeneity of it depending on individual characteristics, such as social background or gender. In the last two decades, a vast number of studies have consistently shown robust evidence on the effect of specific input elements such as class size, teacher quality, instruction time and remedial education in various contexts and across various grade levels.

In the light of the growing consensus on the impact of the core elements on student achievement, some commentators encourage shifting efforts towards an evaluation of the effectiveness of single programmes (Lavy and Schlosser, 2005; Jacob and Lefgren, 2004). Nevertheless, existing evidence on public policies in education most of the time yields mixed results, even when such policies act on input factors the effectiveness of which is commonly acknowledged. In addition, another important element which is receiving considerable attention is the heterogeneity of the effects of these input factors. For example, the huge body of literature on gender achievement gaps still debates feasible ways to reduce the gap in mathematics between males and females, even though this task is made difficult by a lack of a clear diagnosis of the causes of this divide (Ayalon and Livneh, 2012).

This paper focuses on an intervention implemented in 2010 in low-achieving lower secondary schools in four southern Italian regions eligible to receive EU Regional Development Funds and European Social Funds. The Quality and Merit Project[1] (PQM) acts on the quantity of schooling received by pupils, since it mandates more hours of mathematics and Italian language instruction.

School participation is not compulsory, and applicant schools are ranked and selected in order to favour schools with higher rates of dropouts and failing and repeating students. Schools selected to participate must provide extra education (scheduled in the afternoon, outside regular school hours) in mathematics and/or Italian language in two sixth grade classes chosen by the school principal at the time of the application. The extra education comes in the form of both remedial education for the low-achieving students and activities to strengthen the skills of the best students.

Our identification rests upon a difference-in-differences strategy, which compares the test scores of two consecutive cohorts of sixth grade students enrolled in PQM and control classes before and after the implementation of the programme. Sorting of students and teachers across classes is modelled using features of the Italian school system that regulate class composition, which we argue is stable

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

over time. In addition, we investigate possible differences that can emerge according to the gender of the exposed students, and to better understand the hypothesized causal pathways we estimate the effect of the programme on boys and girls' attitudes and self-reported study and classroom behaviours.

Our main results show that additional instruction time in mathematics raises test scores in mathematics for both boys and girls, while no effect of additional instruction time in language is found on language test scores. There are two interesting cross-subject effects which differ by gender: girls who receive more instruction time in mathematics increase their positive attitude toward that subject and also perform better in Italian language, while boys who receive more instruction time in language lower their performance and their positive attitude towards mathematics. We hypothesize that extra language time for boys subtracts time from the study of mathematics, resulting in a lower commitment to this subject, which leads to a decrease in both test scores and attitudes. In other words, extra time in language at school for boys acts as a substitute for engagement in mathematics at home. However, girls attending extra classes in mathematics not only increase their performance in mathematics (and attitude toward it) but also increase their performance in Italian language. We hypothesize that extra instruction time in mathematics provided at school has a double positive effect for girls: it increases the return on the time spent learning mathematics and leaves them with more time at home to study language. In this case, extra classes in mathematics act as a complement with respect to language study time. It is possible that these results can be interpreted by the existing literature studying gender differences in attitude, motivation and discipline.

The rest of the paper is structured as follows: in section 2 we review previous findings about the effect of extra education on student outcomes and part of the literature linking gender difference to achievement; in section 3 we briefly describe the context where the PQM programme was implemented and its structure. Sections 4 and 5 are devoted to the data used and the identification strategy, and in section 6 we present the results. Finally, in section 7 we draw some conclusions and discuss the policy implications resulting from our analysis.

2 Related literature

This section briefly reviews evidence coming from experimental and quasi-experimental research that assesses the effectiveness in improving achievement of specific programmes aimed at adding extra instruction time. The idea behind extra teaching time lies in the simple consideration that the more the student is exposed to school time, the more he/she will learn in a cumulative process. Extra education is also generally conceived to have other side-benefits: it decreases the influence of the family in the case of students of low socio-economic background, and it decreases the negative influence of peers in the case of students exposed to behavioural risks (criminality, teen pregnancy, bullying, etc.). However, assessing the effect of instruction time on student achievement is not an easy task, since extra school time is likely to be correlated with other school resources, as well as the family background of the students. For this reason, much of the evidence on the effect of instruction time on achievement relies on correlational studies, the results of which do not allow for causal inference, at least until studies in the last decade. To overcome this problem, three main strategies have been pursued.

A first strategy makes use of international datasets such as IEA-TIMMS and OECD-PISA to explore the between-country variation in total instruction time per year and the relationship of it with achievement. As Lavy (2010) points out, school systems vary widely with respect to the amount of time students are exposed to different subjects. The first attempts to study the impact of instruction time on performances yielded no significant results (Wößmann, 2003): the effect of instruction time is not of substantial importance once institutional features of each national school system and family characteristics are taken into account. Using OECD-PISA 2006 data and Israeli achievement data on 5th and 8th graders, Lavy (2010) exploits the fact that students are tested in three different subjects (reading, maths and sciences) to perform a within-student estimation of the effect of instruction time on achievement, finding a small positive effect. Using the same approach, Rivkin and Schiman (2013) find similar results focusing on PISA 2009. Moreover, these findings have recently been corroborated by Mandel and Süssmuth (2011) for Germany.

Other scholars rely on quasi-experimental settings exploiting exogenous variation in school year length. Marcotte (2007) and Marcotte and Hemelt (2008) exploit variability in school closure days resulting from snowfalls in Maryland and show that students perform better in the years with fewer unscheduled closure days. Hansen (2008) confirms the existence of a school-year-length effect both using weather-related school-day cancellations (in Colorado and Maryland) and changes in test administration dates that occurred in Minnesota five times in the period 2000-2005. The results confirm the existence of a positive effect of the number of school days on student performance, across various grades and at different points in the performance distribution. Sims (2008) uses a similar idea exploiting a reform in Wisconsin, finding that additional school time is associated with a small increase in performance in mathematics for fourth grade students, but does not affect reading competencies. Using the same strategy (variation in timing of assessment), Fitzpatrick et al. (2011) find a positive effect of schooling days on achievement in both reading and maths for children in Kindergarten and first grade.

A different stream of literature investigates the effect of more school time, but conceived as more hours per day at school rather than more school days per year. Thus, extra education is organized by opening schools longer during the afternoon, either to provide extra instruction time for curricular activities or to help students do their homework. Lavy and Schlosser (2005) estimate the effect of a programme providing targeted additional instruction time to low-achieving high school students in Israel, resulting in an increase in the matriculation rate of about 3%. Bellei (2009) evaluates the impact of the Chilean full school day programme, concluding that the extended school time seems to have been beneficial both for reading and for maths. Zimmer et al. (2010) reports the case of Pittsburgh Public Schools, which enacted various initiatives to improve student performance via extra education and tutoring: the authors were able to identify a positive effect of the two types of programmes in maths but not in reading. More recently, a study conducted by Lavy (2012) on a financial policy reform in Israel provides robust evidence of the effect of the length of the school week on student achievement. Jensen (2013) exploits a policy in Denmark that increased classroom hours in literacy and maths between 2.2 and 3.3 percent and documents very large returns in mathematics, but no effect on reading. A full school day compared to a half school day has also been found to have a positive effect on learning outcomes in kindergarten (Robin et al., 2006; DeCicca, 2007; Lash et al., 2008).

Other programmes, similar in the total number of additional hours and the subject matter, however, have been 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 (2013).

Most of these studies either do not explore gender differences in outcomes or, when they do, they do not find particular differences between boys and girls. In addition, we notice an excess of attention paid to achievement measures at the expense of other potential student outcomes, such as attitudes towards the discipline or study habits. Clearly, achievement is the ultimate goal for which almost all programmes are designed and implemented. However, study of the effects of educational interventions on these alternative outcomes is crucial because attitudes towards school and specific disciplines are strongly connected with achievement (Stevens and Slavin, 1995) and, moreover, it can be helpful to understand which mechanisms are activated to explain the success or failure of a programme. Among all the studies relating time at school to achievement, just one recent paper by Lavy (2012) studies the effect on attitudes and behaviours alongside effects on achievement. Attitudinal effects are studied in terms of school satisfaction and violent behaviour. Nevertheless, he does not find any affect of length of the school week on any of these behavioural measures.

Finally, since our paper also focuses on gender differences both in terms of achievement and in terms of behaviour as a reaction to the provision of extra instruction time, it is worth mentioning a few studies that try to explain gender differences in achievement through differences in learning styles, attitudes, school motivation, and other personality trait differences between boys and girls.

Some studies have suggested that higher motivation and expectations in girls (girls are more serious, diligent and studious, show increased maturity and have more effective learning strategies) are enough to explain their success at school (Fortin et al., 2013) and that boys are more likely to show socially disruptive behaviour, less self-discipline, and anti-school attitudes, which negatively affects their school motivation and educational achievement (Cornwell et al., 2013; Spinath et al., 2010; Steinmayr and Spinath, 2008). In addition, there are gender disparities in both self-perceived ability and interest in maths and reading. Boys are typically more self-confident about their maths skills and show a stronger interest in maths, while for girls the same thing applies to languages (Jacobs et al., 2002). Finally, Golsteyn and Schils (2014) show that gender differences in resources with respect to social and instrumental skills and need for achievement explain part of the differences in performance: boys seem to be better equipped with these resources and boys and girls employ their skills differently.

Our paper can contribute to the existing literature in at least two ways: we provide evidence of the effect of extra instruction time not only on academic achievement, but also on students' attitudes and motivation, and we explore how these effects can differ by gender.

3 Background

Recent international surveys (IEA-PIRLS 2006; IEA TIMMS 2007; PISA 2003, 2006 and 2009) have identified a gap between the Italian school system and those of other OECD countries: 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. A recent experience with national assessment tests has demonstrated that, while the north/south divide is contained for second graders, it increases at the end of primary school and grows even larger in middle schools (INVALSI, 2010). For these reasons, four regions located in the Objective 1 area (Campania, Sicily, Calabria and Apulia) were eligible to benefit from the 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 funds was the implementation of the PQM programme.

The PQM programme targets lower secondary schools (grades six to eight) in the four regions eligible for the PON funding.[2] At the beginning of sixth grade, students are assigned to a specific class, which is called a *sezione*, and they remain in the same class for the whole length of the lower secondary school (i.e. 3 years). This implies that once a student is assigned to a class (*sezione*) he will

[2]The school programmes taught in lower secondary schools are decided by the Italian Ministry of Education and hence are identical across the whole country.

follow all the subjects with the same peers for all three years of the block.[3] In theory, assignment of both teachers and students to the different *sezioni* should be random, but in practice there can be some mechanisms (parental 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. Nevertheless it has always been quite common that a teacher is assigned to the same *sezione* throughout the years and across the grades, and this procedure has been consolidated through a reform that regulates staff deployment, implemented in 2009, explicitly suggesting that the same teacher should be employed in the same *sezione* across the three years foreseen in lower secondary school.[4] Moreover, suggested criteria by the Ministry of Education and common practice in the Italin system is to allocate students so to guarantee homogeneity between classes and heterogeneity within a class.

The programme was first implemented in the academic year 2009/10, subsidizing additional hours in mathematics in 310 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 had already participated in the previous year. In both rounds, participation was not compulsory: applicant schools were enrolled giving preference to worse performing ones according to the percentage of failing and repeating students and school dropout rates. The criteria used for admission were the same in both years. The selection process was intended to favour the more disadvantaged schools among the applicants.

Schools apply to participate in PQM in June, and are notified of acceptance by the end of August. Thus both application and admission take place a few months after the parents have already enrolled their children in a given school, and after the children have been assigned to a specific class (*sezione*), since enrolment of children into lower secondary schools happens between January and February. Thus, it is not likely that parents decided to enrol their children in a given school and exerted pressure to have their children in a given class conditional on the PQM programme.

Participating schools organize 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 choose the two teachers who will provide the extra education, and thus the corresponding two classes that will be treated.[5]

[3]To provide an example, assume that a given lower secondary school is composed of a total of 6 classes: 2 sixth grade classes, 2 seventh grade classes and 2 eighth 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 who is assigned to *sezione* A in the 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.

[4]See section 5 for details

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

Teachers are identified in the application since a part of the intervention foresees that the teachers of the selected classes undertake a training course, the aim of which is to help them organize the extra activities that they will hold in the afternoon. The course consists of 60 hours (30 hours of formal training and 30 hours online) and it helps the teachers to set up an *Improvement Plan*, based on the results of the standardized test which the treated classes take at the beginning of the academic year (October). This test should help teachers target pupils who are in need and areas in which to intervene. The training course is held in groups of 10 teachers (i.e. 5 schools), and it is supervised by a mentor, who provides support regarding their decisions about how to organize the 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 or their knowledge of their subjects, but simply supports them in their decisions on how to organize the extra activities and it provides them with some material that can be used during these activities.

There can be from 1 to 8 afternoon activities planned per class, and teachers receive extra salary for their extra loads.[6] Each activity foresees an average of 15 hours of extra education to be held outside the regular school time, and the teacher is free to decide how many activities and how many students to involve. Activities can be either remedial, thus targeting low achievers, or to deepen the knowledge of certain topics, thus also targeting top-performing students. In our data, the average number of students involved as a proportion of class size varies between 25 percent and 100 percent; in more than 75 percent of classes at least 50 percent of the 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 30 to 60 additional hours spent at school by participating students over the school year.

4 Data, selection of the relevant sample and descriptive statistics

4.1 Data

Data at the school level are provided by the Italian Ministry of Education, through INVALSI. This administrative data provides general information about school characteristics (number of students, student-to-teacher ratio, dropout rates, ...) and the municipality where the school is located. Information on the geographical and demographic characteristics of the environment where the schools operate is also available.

Data at the student level are collected directly by the INVALSI, which is in charge of testing Italian

[6]Teachers receive 50 euro per hour gross. Thus, considering their salary, planning 4 activities would make one months salary.

students' performances through a national assessment test in mathematics and language. This test was introduced in second and fifth grade in a small sample of schools in the academic year 2007/08, and since the academic year 2009/10 it has been taken by all students in the country at the end of second, fifth, sixth, and eighth grades. The language tests are designed to measure reading proficiency (in particular, the ability of students to understand and interpret a text) and lexical and grammatical knowledge, while the mathematics test measures knowledge of mathematical contents and the logical and cognitive processes used in mathematical reasoning. The tests are composed mainly of multiple choice questions in which the students have to select the right answer out of two or four possibilities; in mathematics there are also a few open questions. The score provided by the INVALSI is calculated simply as a percentage of correct answers out of 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 100 .

The data contain information on the results of the standardized tests, both for mathematics and language, the main socio-demographic characteristics of the child and his/her family (gender, year of birth, origin, level of education and employment status of the parents, household composition). A part of the questionnaire is dedicated to students' perceptions, motivation, attitudes and study habits. The outcomes we consider are attitudes towards mathematics and language, study habits, motivation and anxiety during tests, and school satisfaction. Attitudes towards mathematics and language are measured using 4-point-scale questions (ranging from *I do not agree at all* to *I totally agree*) for a total of 4 items for mathematics and 4 items for language, which we summarize with one factor capturing attitudes toward mathematics and one factor capturing attitudes toward Italian language[7]. Study habits are measured via 15 questions reporting the frequency (*Never/Seldom/Often/Always*) of specific actions and techniques of study. In particular, this battery of items refers to organizational ability in studying and cognitive strategies applied at school. Through a factor analysis we identify two main factors, one capturing ability to remember and link useful information both in class and while doing homework (which we name "Ability to link and remember information"), and the other one capturing a tendency to repeat things while studying (which we name "Repeating while studying")[8]. Motivation is measured by 7 items, which through a factor analysis we summarize into two main factors, one measuring "External motivation", which captures a tendency to study and perform well in order to please parents or teachers, and the other one measuring "Internal motivation", which captures a

[7]The items deal with affective as well as cognitive feeling towards the discipline. Two factors are extracted through two-factor analysis measuring attitudes toward mathematics and language. The original number of items was 5 for each subject, but in both cases reliability was higher dropping one of the items. Thus each of the factors is based on 4 items.

[8]The first factor is composed of 7 items and the second of 3 items. From the factor analysis a third factor was extracted, but its reliability was very low. Thus we decided to consider just the first two factors.

tendency to study and perform well for personal satisfaction. School satisfaction is measured through a factor coming from 5 items asking about level of satisfaction with the school based on a 4-point scale. Finally, the questionnaire contains a battery related to test anxiety, measuring agreement (4-point scale, as before) with 7 items, which we summarise into one factor.

In addition, we have information about the *sezione* in which each student is enrolled; this information, as previously explained, is crucial to control for differential sorting of students and teachers into the different classes.

4.2 Selection of the relevant sample and descriptive statistics

In order to evaluate the effectiveness of PQM, two preliminary steps are required: the choice of the group of PQM schools and the choice of control schools. As far as the choice of treated schools is concerned, we focus only on the *second* wave of PQM, which was implemented in the year 2010/11. This choice is driven by data availability: 2009/10 is the first school year in which the national test was compulsory for all sixth grade classes in the country. In other words, we do not have pre-treatment data for the first wave of PQM. Thus we 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 wave. We also decided to drop from the analysis schools which participated in the programme in both years, thus concentrating just on schools which were selected for the first time in 2010/11. In addition, we drop from the sample schools which were enrolled in the programme in the first year but not in the second, since they could not be properly considered control schools.

Table 1 shows descriptive statistics for all the schools in the four regions. The number of PQM schools varies between the four regions[9], with 10 schools in Calabria and around 40 in the other three regions. The tables reveal a differential process of selection of schools in each region. In Campania, the two groups of schools are not dissimilar over a wide range of variables, except for the student-to-teacher ratio and the school size: PQM schools are bigger and have a higher student-to-teacher ratio. In Apulia, PQM schools are bigger and perform worse than non-PQM schools: the percentage of correct answers in PQM schools is on average 2 percentage points lower in both mathematics and Italian language; PQM schools, moreover, show a higher rate of disabled students than non-PQM schools. In Calabria, PQM schools have a higher proportion of permanent teachers and a higher student-to-teacher ratio, and PQM schools are bigger than non-PQM ones and are located in larger towns. In addition, a lower proportion of students attend more than 30 hours per week. Finally, in Sicily the only two significant differences regard student activities and the location of the schools.

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

What seems to be common among the four regions is school size: PQM schools are bigger than non-PQM schools, probably because of the requirements (there 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 choose PQM schools (retention and dropout rates) differentiate between the selected and non-selected schools.

[Insert Table 1 here]

In order to control for bias resulting from the non-random selection of the school, we choose a group of control schools sharing similar observable characteristics to the schools enrolled in the programme through propensity score matching. The propensity score is calculated separately in each region, and the matching is done one-to-one with replacement inside the same province. The matching procedure along the dimension considered does not yield any common support problem.[10] The variables used to estimate the propensity score are: average percentages of correct answers in mathematics and language in sixth grade in 2009/10, student-teacher ratio, proportion of permanent teachers, dropout rate, fail rate, proportion of repeating students, proportion of immigrant students, proportion of disabled students, proportion of female students, proportion of students attending more than 30 hours per week, total number of students, use of PON funding in 2009/10, population of the town and whether the school is located in a mountain municipality. The results of the logistic regressions are shown in Table 2.

[Insert Table 2 here]

Once the propensity score is obtained, we match each PQM school with the non-treated school located in the same province with the closest propensity score. Table 3 shows the final numbers of schools, classes and students in PQM and control schools, in both the pre- and post-treatment years. The final sample of treated schools is composed of 23 schools only enrolled in PQM mathematics, 37 schools only enrolled in PQM Italian language and 74 schools enrolled in both disciplines. This corresponds to 127 classes receiving extra education in mathematics, 146 in Italian language and 40 in both subjects during the academic year 2010/11. [11]

[Insert Table 3 here]

[10]Only 4 PQM schools out of 138 had to be dropped because of this.

[11]Although the number of schools selected for both programmes was high (74 schools out of 134), only a few classes were selected to participate in both parts of the programme. This happened in smaller schools, where it was impossible to implement the programme in 4 different classes.

Table 4 reports some relevant descriptive statistics for the PQM schools and the schools chosen as controls. The average of the various dimensions considered is similar see columns (1) and (2) and not statistically different between the two groups see column (3). In column (4) the estimates of a logistic regression for the probability of being a PQM school in the working sample are reported. It turns out that, after 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. Table 5 presents descriptive statistics of average student 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 individual characteristics: pupils in control schools have a higher percentage of mothers working and a lower percentage of unemployed fathers.

[Insert Table 4 here]

[Insert Table 5 here]

5 Method

With the matching procedure, we select a control group of schools sharing similar observable characteristics. In order to take into account unobservable characteristics and to provide an unbiased estimate of the effect of the programme, we apply a difference-in-differences approach.

Using school identifiers provided by the INVALSI, we were able to link data for the same school in 2009/10 (pre-programme period) and 2010/11 (post-programme period). Moreover, we obtained identifiers for the *sezione* to which students are assigned at school. The treatment status is defined at the class level, and for estimation purposes we use data on the two cohorts of sixth graders in 2009/10 and 2010/11. In practice, we compare performances of two contiguous cohorts of children belonging to the same *sezione* at the end of the sixth grade, before and after the programme implementation. This is a standard difference-in-differences approach, with *sezione* fixed effects. The key feature that we exploit to control for endogenous sorting of students and teachers over classes is that it is common practice in Italian schools to maintain the same teachers in the same *sezione* over the years and across grades. Adoption of this practice is prompted by a reform that was implemented nationwide in 2009 to regulate teaching ratios [12]. For example, the law states that maths teachers must fulfil their weekly duties at school by teaching modules of 6 hours to three different classes, explicitly suggesting that these should be the sixth, seventh and eighth grades of the same *sezione*. In this context, we argue

[12] Decreto ministeriale number 37, 26 March 2009

that controlling for *sezione* fixed effects is roughly like controlling for teacher fixed effects.[13] Note that by controlling for *sezione* fixed effects we indirectly control for school fixed effects, and thus for sources of potential bias related to unobservable characteristics of the classes and of the schools.

Finally, we are not able to identify the students who really participate in the afternoon activities. Therefore, although we are able to detect the effect of belonging to a class enrolled in the PQM, we cannot see the marginal effect of one more hour of schooling in a given subject.

Our basic specification considers the following equation:

$$y^k_{ijt} = \beta_0 + \beta_1 C^M_{jt} T_{jt} + \beta_2 C^L_{jt} T_{jt} + \beta_3 N_{jt} T_{jt} + \beta_4 T_j + \theta X_{ijt} + \gamma_j + \epsilon^k_{ijt}, \tag{1}$$

where y^k_{ijt} is the percentage of correct answers in subject k (mathematics or language) for student i in *sezione* j in year t. T_{jt} is an indicator of observations in the post-intervention year. C^L_{jt} and C^M_{jt} are dummies for being enrolled in *any* activity in Italian language (L) and mathematics (M) respectively, while N_{jt} is a dummy for control classes in PQM schools. X_{ijt} is a vector of student and class characteristics, γ_j is the *sezione* fixed effect and ϵ^k_{jt} is a random error. The coefficients β_1 and β_2 are the main effects of interest, and β_3 captures possible spillover effects of treated classes on non-treated classes in PQM schools. All standard errors are clustered at the school level.

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

This general model form is also applied to continuous factors extracted from the questionnaire that we use to measure students' attitudes, study behaviour and motivation. In order to assess whether the intervention has different effects for males and females, all the analyses are replicated separately for males and females.

6 Results

Table 6 shows the estimated effects of the intervention on the percentage of correct answers in mathematics and Italian language. All the models shown hereafter control for class size, number of regular school hours per week, gender, migration status, regularity status (whether a student is ahead or behind compared to his age), maximum level of education of the parents, mother working status and

[13] We rely on the same identification strategy used in Battistin and Meroni (2013).

dummies indicating missing data (individual-level variables only; the class-level variables show no missings).

We present the results for the whole sample of students (column (1) for the mathematics score and column (4) for the Italian language score) and also for the sub-samples divided by gender. The results for females are shown in columns (2) and (5) and for males in columns (3) and (6). Considering the whole sample, we see a positive effect of extra time at school in mathematics on mathematics test scores, with the intervention increasing the percentage of correct answers by 2.4. On the other hand, no effect is found for extra time in language on Italian language test scores. Other studies dealing with the effect of instruction time have documented similar findings (Sims, 2008; Jensen, 2013; Zimmer et al., 2010).

As for spillover effects inside the schools, we find none: the coefficients associated with being in a control class in a PQM school are neither substantially nor significantly different from 0. Thus, control classes in PQM schools are not different from classes in control schools. In addition, there are no cross-subject effects: receiving extra time in mathematics does not have an effect on language outcomes, and vice versa.

However, if we split the sample and look at the results by gender, we find a different and rather surprising picture. For both sub-groups we notice a persistence of the positive effect of receiving extra time in mathematics on the mathematics test score, with the percentage of correct answers increasing by 2.7 points for females and by 2.3 points for males, and no effect of receiving extra time in language on the language test score. Nevertheless, we also notice two interesting cross-subject effects: for females, a positive effect of receiving extra time in mathematics on their Italian language outcome; and for males, a negative effect of receiving extra time in language on their mathematics outcome. The cross-subject effects are smaller than the direct effects and significant only at the 10% level, but it still emerges that girls receiving extra classes in mathematics also perform better in language, and that boys receiving extra classes in language perform worse in mathematics.

[Insert Table 6 here]

As a further check, we investigate whether the intervention could reduce the gap in performances that exists between boys and girls (with boys out-performing girls in mathematics and girls out-performing boys in language). Therefore, in each class we calculate the ratio of the percentage of correct answers given by boys to the percentage of correct answers given by girls in the two subject tests, and then estimate whether these ratios are changed by spending more time at school. The result is that neither of the two ratios is affected by the PQM programme. Thus, the intervention does not significantly discriminate between boys and girls: both increase their test scores in mathematics, but the programme does not manage to close the gaps between girls and boys in either mathematics or Italian language[14].

In order to better understand the results for achievement, we also estimate the effect of spending more time at school on the variables capturing attitudes, motivation, study behaviour and school satisfaction.

Table 7 presents the results. We find that extra time in language increases the factor associated with a positive attitude towards that subject for both boys and girls, and decreases the factor associated with positive attitudes towards mathematics for boys. On the other hand, mathematics time increases the factor associated with positive attitudes towards mathematics only for girls. We find that neither internal nor external motivation are affected by the programme, and neither are school satisfaction or study behaviours (only girls who attend more classes in Italian show a slightly higher score for the ability to link and remember information). These results are in line with the only assessment of the impact of extra time on student behaviour and satisfaction that we are aware of (Lavy, 2012). Finally, boys attending more classes in language seem to be more anxious during tests.

If we try to link the results for achievement and those for the psychological variables, we notice that boys receiving extra time in language lower their performance in mathematics and also show a decrease in the factor associated with a positive attitude toward that subject. It is not possible with our data and it is not our purpose to disentangle the effect of attitudes on achievement and/or vice versa. Nevertheless, the analysis clearly shows that receiving extra school hours in language negatively affects both performance and attitude toward mathematics for boys. We can hypothesize that extra time in language, for boys, subtracts time from the study of mathematics resulting in a lower commitment to this subject. In other words, for boys, receiving extra instruction time in Italian language can act as a substitute for engagement in maths. This result is in line with observations in the literature that boys tend to be less motivated and have less discipline than girls (Steinmayr and Spinath, 2008; Spinath et al., 2010; Cornwell et al., 2013; Fortin et al., 2013). Thus, we can hypothesize that spending more time at school doing Italian language comes at the cost of reducing the time boys spend studying at home, which leads to a reduction in performances in mathematics. In addition, this time spent at school studying language is useful only in increasing their positive attitude toward that subject, but is not enough to lead to an increase in academic performance.

As far as girls are concerned, we observe the reverse situation: girls who receive more instruction time in mathematics show a significant increase in maths scores and a small increase in language

[14] Table available upon request

performance. We hypothesize that attending extra classes in mathematics (where females normally have greater difficulties than males) helps them not only improve their mathematics achievement and their positive attitude towards it, but also helps increase their performance in Italian language. It is possible that extra instruction time in mathematics at school increases the return on the time they spend studying mathematics and leaves them with more time at home to study language, also leading to an increase in their Italian language test scores. In this case, extra classes in maths act as a complement to language study time. Again, this result is consonant with findings in the existing literature that girls are characterized by greater motivation and discipline and more risk aversion (Steinmayr and Spinath, 2008; Borghans et al., 2009; Spinath et al., 2010; Cornwell et al., 2013; Fortin et al., 2013).

This cross-disciplinary effect for females of receiving more instruction time in mathematics on their language test score where the language programme has proven to be ineffective may be surprising at first glance, but it can be explained by examining it within a gender dimension. If receiving extra instruction time in mathematics helps girls gain confidence in that subject, where they are usually behind, it can also lead to spillover effects on Italian language: girls are motivated and committed, and seeing the results they are obtaining in mathematics can push them to invest more and thus perform better in Italian language too.

[Insert Table 7 here]

7 Conclusion

In this paper we have examined an EU-funded intervention providing extra school time in mathematics and language to sixth grade students. The intervention targeted students located in the most disadvantaged area of the country and it aimed to increase performances in the two subjects. In line with many other studies of this kind (Sims, 2008; Jensen, 2013; Zimmer et al., 2010), it seems to be easier for an increase in instruction time to improve achievement in mathematics than in Italian language: students receiving extra instruction time in mathematics perform better in that subject and develop a more positive attitude towards the discipline, while no effect is found for students receiving extra instruction time in Italian language.

If we take a closer look at the gender dimension, however, we can see different mechanisms at play among male and female students. While involvement in the programme seem to have been very effective for females, for boys we find controversial results. Girls receiving extra instruction time in mathematics improve their achievement both in mathematics and in language; on the other hand, boys receiving extra instruction time in mathematics increase their performance in mathematics, but extra instruction time in language worsens their performance in mathematics.

We interpret these results in the context of the literature studying gender differences in achievement and in attitudes toward school commitment. The stimuli provided by extra classes are received differently by male and female students. We hypothesize that the programme may have acted as a substitute for normal study time for boys, but as a complement to it for girls. A reason for this could lie in the higher risk-aversion, motivation, commitment and discipline that girls have with respect to boys.

Future research should explore the heterogeneity of educational policy effects in more depth: among other characteristics, gender is often neglected but it seems to be a crucial dimension to consider to target educational interventions better.

References

Ayalon, H. and I. Livneh (2012). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social science research 42*, 559–574.

Battistin, E. and E. C. Meroni (2013). Should we increase instruction time in low achieving schools? evidence from southern italy. Technical report, IZA Discussion Paper.

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

Borghans, L., J. J. Heckman, B. H. Golsteyn, and H. Meijers (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association 7*(2-3), 649–658.

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

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

Cornwell, C., D. B. Mustard, and J. Van Parys (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources 48*(1), 236–264.

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

Dynarski, M., S. James-Burdumy, M. Moore, L. Rosenberg, J. Deke, W. Mansfield, et al. (2004). 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*.

Fitzpatrick, M., D. Grissmer, and S. Hastedt (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review 30*(2), 269–279.

Fortin, N. M., P. Oreopoulos, and S. Phipps (2013). Leaving boys behind: Gender disparities in high academic achievement. Technical report, National Bureau of Economic Research.

Golsteyn, B. and T. Schils (2014). Gender gaps in primary school achievement: A decomposition into endowments and returns to iq and non-cognitive factors. *Economics of Education Review, Forthcoming.*

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

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

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

Jacobs, J. E., S. Lanza, D. W. Osgood, J. S. Eccles, and A. Wigfield (2002). Changes in childrens self-competence and values: Gender and domain differences across grades one through twelve. *Child development 73*(2), 509–527.

Jensen, V. M. (2013). Working longer makes students stronger? the effects of ninth grade classroom hours on ninth grade student performance. *Educational Research 55*(2), 180–194.

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

Lavy, V. (2010). 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*.

Lavy, V. (2012). 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*.

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

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

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

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

Meyer, E. and C. Van Klaveren (2013). The effectiveness of extended day programs: Evidence from a randomized field experiment in the netherlands. *Economics of Education Review 36*, 1–11.

Rivkin, S. G. and J. C. Schiman (2013). Instruction time, classroom quality, and academic achievement. Technical report, National Bureau of Economic Research.

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

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

Spinath, B., H. Harald Freudenthaler, and A. C. Neubauer (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences 48*(4), 481–486.

Steinmayr, R. and B. Spinath (2008). Sex differences in school achievement: what are the roles of personality and achievement motivation? *European Journal of Personality 22*(3), 185–209.

Stevens, R. J. and R. E. Slavin (1995). The cooperative elementary school: Effects on students achievement, attitudes, and social relations. *American Educational Research Journal 32*(2), 321–351.

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

Zimmer, R., L. Hamilton, and R. Christina (2010). 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.

Tables and Figures

Table 1: Descriptive statistics for PQM and non PQM schools,

	Campania			Apulia			Sicily			Calabria		
	PQM	Non PQM	Diff.	PQM	Non PQM	Diff.	PQM	Non PQM	Diff.	PQM	Non PQM	Diff.
Mathematics, percentage of correct answers	0.578	0.582	-0.004	0.578	0.597	-0.019	0.546	0.554	-0.008	0.596	0.560	0.037
			(0.010)			(0.010)			(0.011)			(0.020)
Italian, percentage of correct answers	0.489	0.486	0.003	0.478	0.503	-0.025	0.457	0.467	-0.010	0.503	0.479	0.024
			(0.014)			(0.011)			(0.015)			(0.029)
Proportion of permanent teachers	0.917	0.890	0.027	0.887	0.901	-0.014	0.855	0.843	0.012	0.924	0.814	0.110
			(0.016)			(0.017)			(0.021)			(0.047)
Student-teacher ratio	9.933	9.038	0.895	10.34	10.31	0.032	8.655	8.594	0.061	9.409	7.486	1.923
			(0.380)			(0.339)			(0.061)			(0.734)
Number of students in the school	430.3	334.7	95.54	438.9	349.5	89.43	356.3	299.2	57.10	380.4	221.7	158.7
			(35.89)			(32.10)			(30.05)			(51.35)
Proportion of immigrant students	0.022	0.025	-0.002	0.028	0.027	0.001	0.029	0.034	-0.005	0.035	0.037	-0.002
			(0.004)			(0.005)			(0.009)			(0.012)
Proportion of disable students	0.031	0.034	-0.003	0.032	0.026	0.005	0.044	0.040	0.004	0.025	0.030	-0.004
			(0.004)			(0.003)			(0.005)			(0.006)
School drop out rate	0.001	0.002	-0.000	0.003	0.002	0.001	0.005	0.004	0.001	0.005	0.004	0.002
			(0.001)			(0.001)			(0.002)			(0.004)
School rate of failing students	0.041	0.040	0.001	0.038	0.033	0.006	0.069	0.065	0.004	0.037	0.045	-0.007
			(0.007)			(0.006)			(0.001)			(0.015)
School rate of repeating students	0.038	0.037	0.002	0.040	0.032	0.008	0.064	0.071	-0.007	0.051	0.038	0.013
			(0.007)			(0.005)			(0.010)			(0.013)
Proportion of female in the school	0.492	0.478	0.0148	0.480	0.484	-0.004	0.504	0.484	0.020	0.477	0.476	0.001
			(0.012)			(0.010)			(0.012)			(0.027)
Proportion of students more than 30 hours	0.394	0.384	0.009	0.262	0.314	-0.052	0.371	0.407	-0.035	0.242	0.527	-0.285
			(0.068)			(0.063)			(0.070)			(0.023)
School received PON funds	0.930	0.859	0.072	1.000	0.960	0.040	0.976	0.874	0.103	0.900	0.877	0.023
			(0.055)			(0.030)			(0.052)			(0.106)
Municipality located on mountain	0.233	0.309	-0.076	0.326	0.186	0.140	0.286	0.449	-0.163	0.400	0.626	-0.226
			(0.073)			(0.066)			(0.080)			(0.157)
(Log) Population in town	10.12	9.961	0.159	10.61	9.969	0.644	10.51	10.16	0.356	10.47	9.016	1.457
			(0.298)			(0.221)			(0.280)			(0.431)
Number of schools	43	460		43	253		42	408		10	227	

Note. Presented are the descriptive statistics of PQM and non PQM schools and the difference, with corresponding standard errors in the four regions.

Table 2: Probability of being a PQM School

	Campania	Apulia	Calabria	Sicily
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

Note. Presented 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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Sample size (schools, classes and students)

	Pre treatment year	Post treatment year
Schools		
Enrolled in PQM	134	134
Used as control	114	114
Classes		
Enrolled in PQM	313	313
Used as controls in PQM schools	407	407
Used as controls in remaining schools	595	595
Students		
Enrolled in PQM	6228	6461
Used as controls in PQM schools	8260	8380
Used as controls in remaining schools	12455	12672

Note. Presented are the number of students, classes and schools in the working sample (see Section 4.2 for details).

Table 4: Descriptive statistics (schools enrolled in PQM and matched control schools) - pre-programme data only

	(1) PQM	(2) Control	(3) Difference	(4) Score
% of correct answers in mathematics	0.480	0.489	-0.009	1.418
			(0.010)	(2.861)
% of correct answers in Italian language	0.572	0.584	-0.012	-4.010
			(0.008)	(3.639)
% of permanent teachers	0.892	0.904	-0.012	-0.341
			(0.012)	(1.612)
Student to teacher ratio	9.632	9.931	-0.299	-0.098
			(0.275)	(0.099)
Number of students	402.8	398.4	4.470	0.001
			(26.34)	(0.001)
% of female students	0.490	0.488	0.002	0.589
			(0.007)	(2.385)
% of foreign students	0.027	0.027	0.000	1.088
			(0.003)	(5.101)
% of students with disabilities	0.034	0.031	0.003	2.652
			(0.003)	(8.669)
% of repeating students	0.048	0.041	0.007	4.105
			(0.006)	(4.182)
Drop out rate	0.003	0.003	0.000	-5.368
			(0.001)	(12.07)
Failure rate	0.049	0.046	0.003	-3.665
			(0.006)	(3.912)
% of classes doing more than 30 hours	0.335	0.337	-0.001	-0.170
			(0.051)	(0.400)
% received PON funds	0.963	0.974	-0.011	-0.273
			(0.023)	(0.769)
Municipality located on mountains	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		

Note. Presented are descriptive statistics for schools in the working sample obtained as described in Section 4.2. Column (1) refers to schools enrolled in the programme; column (2) refers to schools used as controls; column (3) is the difference between column (1) and column (2), and the standard error of the difference in reported in parentheses; column (4) reports the results from a logit regression for being a PQM school (standard error in parentheses).

Table 5: Descriptive statistics (students in schools enrolled and matched control schools) - pre-programme data only.

	(1) PQM	(2) Control	(3) Difference
Test score mathematics	41.66	42.39	-0.728
			(1.015)
Test score language	57.52	58.57	-1.055
			(0.884)
% of correct answers in mathematics	0.483	0.491	-0.008
			(0.010)
% of correct answers in Italian language	0.575	0.585	-0.010
			(0.008)
% of ahead students	0.026	0.030	-0.004
			(0.004)
% of behind students	0.060	0.054	0.005
			(0.006)
Class average weekly hour	31.75	31.42	0.336
			(0.316)
Average class size	21.96	22.43	-0.469
			(-0.469)
% of mothers employed	0.362	0.422	-0.060
			(0.021)
% of students whose father's occupation is: unemployed	0.071	0.052	0.018
			(0.009)
% of students whose father's occupation is: blue collar	0.307	0.304	0.003
			(0.019)
% of students whose father's occupation is: white collar	0.425	0.425	0.000
			(0.017)
% of students whose father's occupation is: managerial	0.197	0.218	-0.021
			(0.016)
% of students whose parents have low education	0.459	0.407	0.052
			(0.027)
% of students whose parents have medium education	0.399	0.415	-0.017
			(0.018)
% of students whose parents have high education	0.142	0.177	-0.035
			(0.018)
Average HOME scale coefficient	-0.065	-0.049	-0.016
			(0.032)
% of students living with both parents	0.900	0.892	0.008
			(0.006)
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	

Note Presented are descriptive statistics obtained as described in Section 4.2. Column (1) refers to average characteristics of students in PQM schools; column (2) refers to average characteristics of students in schools used as controls; column (3) is the difference between column (1) and column (2), and the standard error of the difference in reported in parentheses.

Table 6: Effect of PQM on the percentage of correct answers in mathematics and Italian language

	Mathematics			Italian language		
	(1)	(2)	(3)	(4)	(5)	(6)
	Whole sample	Females	Males	Whole sample	Females	Males
Any extra class in mathematics	0.024**	0.027**	0.023*	0.010	0.015*	0.010
	(0.012)	(0.013)	(0.013)	(0.009)	(0.009)	(0.010)
Any extra class in language	-0.010	0.001	-0.019*	0.002	0.005	-0.002
	(0.010)	(0.012)	(0.011)	(0.008)	(0.009)	(0.010)
Control class in PQM schools	-0.001	-0.002	0.002	-0.007	-0.009	-0.003
	(0.009)	(0.010)	(0.009)	(0.007)	(0.008)	(0.008)
Observations	54456	26315	27783	54456	26315	27783

Note. Difference-in-differences estimates of the effect of the intervention on mathematics (columns (1), (2) and (3)) and Italian language (columns (4), (5) and (6)) percentage of correct answers, on the whole sample of students (column (1) and (4)), on the subsample of females (column (2) and (5)) and on the subsample of males (column (3) and (6)). Each column correspond to a separated regression. Estimates are at the student level with *sezione* fixed effects. Standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Effect of PQM on attitude and behaviuor

	(1)	(2)	(3)	(4)	(5)	(6)			
	Positive attitude towards Italian language			Positive attitude towards mathematics					
	Whole sample	Females	Males	Whole sample	Females	Males			
Any extra class in mathematics	-0.029	-0.048	-0.001	0.040	0.080*	0.019			
	(0.042)	(0.048)	(0.053)	(0.034)	(0.048)	(0.045)			
Any extra class in language	0.133***	0.116**	0.156***	-0.033	0.039	-0.114***			
	(0.039)	(0.046)	(0.051)	(0.033)	(0.047)	(0.043)			
Control class in PQM schools	0.022	-0.002	0.031	0.001	0.021	-0.007			
	(0.031)	(0.037)	(0.037)	(0.025)	(0.035)	(0.034)			
Observations	49704	24073	25330	49992	24204	25488			
	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	External motivation			Internal motivation			Anxiety		
	Whole sample	Females	Males	Whole sample	Females	Males	Whole sample	Females	Males
Any extra class in mathematics	0.029	0.073	0.003	-0.027	-0.017	-0.048	-0.013	-0.002	-0.010
	(0.032)	(0.045)	(0.045)	(0.034)	(0.047)	(0.049)	(0.037)	(0.047)	(0.048)
Any extra class in language	-0.019	0.018	-0.057	0.015	0.010	0.001	0.080**	0.065	0.087*
	(0.029)	(0.044)	(0.040)	(0.034)	(0.043)	(0.047)	(0.034)	(0.045)	(0.045)
Control class in PQM schools	0.056*	0.038	0.090**	0.005	0.016	-0.017	0.019	0.021	0.030
	(0.029)	(0.038)	(0.037)	(0.026)	(0.036)	(0.033)	(0.028)	(0.035)	(0.037)
Observations	46854	23793	22775	46854	22775	23793	50241	24415	25533
	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	Link and remember information			Repeating while studying			Satisfaction		
Any extra class in mathematics	0.041	0.028	0.060	0.023	0.014	0.038	0.037	0.025	0.049
	(0.039)	(0.053)	(0.048)	(0.033)	(0.043)	(0.044)	(0.045)	(0.056)	(0.052)
Any extra class in language	0.057	0.096*	0.013	0.007	0.050	-0.035	0.024	0.004	0.048
	(0.036)	(0.049)	(0.042)	(0.035)	(0.048)	(0.043)	(0.047)	(0.052)	(0.056)
Control class in PQM schools	0.007	-0.002	0.021	-0.011	0.008	-0.019	-0.005	-0.038	0.031
	(0.029)	(0.037)	(0.038)	(0.025)	(0.034)	(0.033)	(0.033)	(0.040)	(0.039)
Observations	44640	21699	22669	50890	24644	25947	49534	24060	25186

Note. Difference-in-differences estimates of the effect of the intervention on students' attitudes and behaviors. Each column corresponds to a separated regression. Estimates are at the student level with *sezione* fixed effects. Standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$