

Education Economics



ISSN: 0964-5292 (Print) 1469-5782 (Online) Journal homepage: http://www.tandfonline.com/loi/cede20

How do students react to longer instruction time? Evidence from Italy

Elena Claudia Meroni & Giovanni Abbiati

To cite this article: Elena Claudia Meroni & Giovanni Abbiati (2016): How do students react to longer instruction time? Evidence from Italy , Education Economics, DOI: <u>10.1080/09645292.2015.1122742</u>

To link to this article: <u>http://dx.doi.org/10.1080/09645292.2015.1122742</u>

9

© 2016 European Union. Published by Taylor & Francis.



View supplementary material 🖸

-	0

Published online: 19 Jan 2016.

_	
С	
L	12 I
~	

Submit your article to this journal \square

Article views: 263



View related articles 🗹



View Crossmark data 🗹

Full Terms & Conditions of access and use can be found at http://www.tandfonline.com/action/journalInformation?journalCode=cede20

∂ OPEN ACCESS

How do students react to longer instruction time? Evidence from $Italy^{\dagger}$

Elena Claudia Meroni^a 回 and Giovanni Abbiati^b 回

^aEuropean Commission, Joint Research Centre, Via Fermi, Ispra, Italy; ^bFBK – IRVAPP, Via Dalla Piccola, Trento, Italy

ABSTRACT

This paper investigates the effects on achievement, study behaviours and attitudes of an intervention providing extra instruction time in language and in mathematics 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 find that an average increase of 25% in instruction time leads to an increase in 0.12 sd in mathematics test score for both females and males, while no effect is found on Italian language test scores. Cross-disciplinary effects seem to suggest that extra-classes in mathematics are beneficial for girls also for language scores. The pattern of results found on attitudes and self-reported study behaviours suggests that girls use the extra instruction time as a complement to regular home study, while boys may use it as a substitute.

ARTICLE HISTORY

Received 30 June 2014 Accepted 15 November 2015

KEYWORDS

Education policy; policy evaluation; instruction time; gender differences

1. Introduction

Dating back to the mid-1960s and to the publication of the Coleman Report (Coleman et al. 1966), the question of how to improve students skills 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 set of school inputs on student outcomes (typically, student achievement in core subjects such as reading, mathematics and sciences) and their heterogeneity on the basis of 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 (Jacob and Lefgren 2004; Lavy and Schlosser 2005). Nevertheless, existing evidence on public policies in education most of the time yields mixed results, even when such policies act on input factors whose effectiveness is commonly acknowledged.

This paper studies the effects of 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 Project¹ (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

The views expressed in this article are those of the authors and should not be attributed to the European Commission.

B Supplemental data for this article can be accessed at 10.1080/09645292.2015.1122742.

© 2016 European Union. Published by Taylor & Francis.

CONTACT Elena Claudia Meroni 😡 elena.meroni@jrc.ec.europa.eu

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http:// creativecommons.org/Licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

favour schools with higher rates of dropouts, 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 instruction time 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 over time.

This paper focuses on the effect of extra instruction time on both students' achievement and behavioural outcomes, estimating possible heterogeneity in effects driven by the gender of the students. The beneficial effects that an increase in instruction time can exercise on achievement are rather clear: the more the time devoted to a subject, the more the student will learn in a cumulative process. However, there could be also side-effects: staying longer at school may protect students from behavioural risks, but it could also provoke negative side-effects such as boredom or school revulsion. Following the idea proposed by Lavy (2012), we estimate the effect of spending more time at schools not only on students' achievement but also on some variables capturing students' motivation, attitude and behaviour.

The type and the intensity of behavioural and achievement responses are likely to be heterogeneous between students according to key characteristics. Among others, gender seem to be the one most called upon. While it is well known that males and females score differently in humanistic and mathematical disciplines, with girls performing better in the former and boys in the latter, we don't know how the provision of extra instruction time could affect differently their performances and behaviours. Girls are more diligent and mature and adopt more effective learning strategies, but boys are more competitive and risk averse, elements that could lead to differential achievement results within gender even when exposed to the same amount of extra instruction time (Joensen and Nielsen 2013). Moreover, negative behavioural reactions, such as boredom or low commitment, are particularly plausible for male students, since they usually have a low commitment to school workloads (Duckworth and Seligman 2006).

The study of PQM can provide some interesting clues on both the topic of effectiveness of an extra-instruction time program on achievement and on gender differences, for at least two reasons: first, because both math and language programs are structured in the same way; second, because they are addressed to the same cohort of students of the same geographical area. Moreover, the program is sufficiently demanding in terms of instruction time as to potentially provoke both an effect on disciplinary competence and to induce side-effects on students' attitudes and behaviour.

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 towards that subject and also perform better in Italian language, while boys who receive more instruction time in language lower their performance and their attitude towards mathematics. We hypothesize that extra language time for boys could subtract 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 may acts as a substitute for engagement in mathematics at home. On the other side, girls attending extra classes in mathematics not only increase their performance in (and attitude towards) mathematics, but they also increase their performance in Italian language. We hypothesize that extra instruction time in mathematics provided at school can have a double positive effect for girls: it can increase the return on the time spent learning mathematics and eventually leave them with more time at home to study language. In this case, extra classes in mathematics may a ct as a complement with respect to language study time.

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, followed by some robustness checks presented in Section 7. Finally, in Section 8 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 on the link between instruction time and achievement. The idea behind extra teaching time lies in the simple consideration that the more the student is exposed to school time, the more s/he 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 from 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.). On the other hand, an excessive amount of schooling time could lead to phenomena of school disaffection. Assessing the effect of instruction time on student achievement or behaviour 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 its relationship with achievement. As Lavy (2010) points out, school systems vary widely with respect to the amount of time students are exposed to different subjects. Small effects of instruction time are found by Lee and Barro (2001), who uses a panel of 59 countries, by Wößmann (2003), who analyses TIMMS data and by Lavy (2010), who combines OECD-PISA 2006 data with Israeli achievement data on 5th and 8th graders and finds slightly higher effects for females. 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) use variability in school closure days resulting from snowfalls in Maryland and show that students perform better in the years with fewer unscheduled closure days. A similar approach is also used by Hansen (2008) and Fitzpatrick, Grissmer, and Hastedt (2011): along with weather-related school-day cancellations, the variation in timing of assessment is also found to be robustly associated with students achievement. Sims (2008) exploiting a reform in Wisconsin finds positive effects on mathematics test scores, but not on language, while 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 mathematics. New evidence from North Carolina and Massachusetts administrative data confirms the relationship between days spent at school and achievement (Aucejo and Romano 2014; Goodman 2014). The authors, however, put forward that much of the effect is connected to individual behaviours, such as absences, rather than to changes in school calendars (e.g. weather-related school schedule modifications). Finally, Parinduri (2014) finds that a longer school year increases educational attainment and decreases grade repetition in Indonesia.

A different stream of literature investigates the effect of additional school time 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%. Zimmer, Hamilton, and Christina (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 mathematics but not in reading. More recently, a study conducted by Lavy (2012) on a financial policy reform in Israel provides robust evidence of the positive 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% and documents very large returns in mathematics, but no effect on reading. Cortes, Goodman, and Nomi (2015) study the effect of receiving algebra courses with double instructional time in ninth grade finding sizable effects on high school graduation rates, college entrance exam scores, and college enrolment rates, especially among low achieving students. Falch, Nyhus, and Strøm (2014) finds that students in Norway who are subject to an intensive preparation period, and examination in mathematics decrease dropout from high school, increase enrolment in higher education, and increase enrolment in natural science and technology education programs, the causal effects being somewhat stronger for males than for females and among highly skilled students.

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, Frede, and Barnett 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). The evidence here reported shows that the picture is very heterogeneous: various programmes proved to be effective only in one specific subject or ineffective at all.

What is lacking in this literature is a discussion of the potential benefits of extra-education time on outcomes other than achievement, such as attitudes and study behaviours. Clearly, achievement is the ultimate goal for which almost all programmes are designed and implemented. However, study of the effects of educational interventions on alternative outcomes is crucial because, for instance, attitudes towards school and/or specific disciplines are strongly connected with achievement (Stevens and Slavin 1995). Among all the studies relating time at school to achievement, just one recent paper by Lavy (2012) studies the effect on non-cognitive outcomes. Attitudinal effects are studied in terms of school satisfaction and violent behaviour: the author does not find any effect of length of the school week on any of these behavioural measures.

Another reason for which a look on how extra-instruction time affects students' perceptions or their school experience can be helpful is to have some clues on which mechanisms are activated to explain the success or failure of a programme. Longer school days can be managed by the students in many ways: there can be a re-allocation of their homework and study time, in order to face new time constraints; they can react positively, gleaning more enthusiasm and motivation towards school or, conversely, generating repulsion towards school-related tasks, such as homework or specific disciplines.

An important domain in which there may be marked differences in this respect is gender. Some evidence shows that girls are more serious, diligent and studious, show increased maturity and have more effective learning strategies (Martin 2004; Duckworth and Seligman 2006; Xu 2006; Wagner, Schober, and Spiel 2008; Golsteyn and Schils 2014). These attitudes and organizational skills for some authors are enough to explain their success at school (Fortin, Oreopoulos, and Phipps 2013), where they normally outperform boys, considering also 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 (Steinmayr and Spinath 2008; Spinath, Harald Freudenthaler, and Neubauer 2010; Cornwell, Mustard, and Van Parys 2013).

Nevertheless, it is widely recognized that boys perform better than girls in mathematics, and the gap in performance increases gradually as we move from the mean to the top of the performance distribution. Explanations for this gender gap in mathematics have been identified in the fact that

females tend to have higher risk aversion and lower preference for competition than men and are less self-confident and more uncertain about their own abilities (Jacobs et al. 2002; Croson and Gneezy 2009). Buser, Niederle, and Oosterbeek (2014) find that the gender gap in choice of academic track, is substantially explained by the gender gap in competitiveness and since mathematics is usually a challenging subject, females often do not perform as high as they could, and choose to focus their career on different subjects (Niederle and Yestrumskas 2008). Girls seem to have particularly little faith in their own math abilities also due to extensive gender-stereotyping regarding math, but a recent study by Joensen and Nielsen (2013) showed that when females were encourage to follow advance mathematics course, had positive returns on future earnings higher than males.

The programme under scrutiny can contribute to the existing literature in at least two ways: first, by providing evidence on the effect of extra instruction time not only on achievement but also on students' attitudes and motivation; second, by documenting gender differences in both achievement and study habits and behaviours when expose to more time dedicated to either mathematics or language.

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 specific (*sezione*) he/she will 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 *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 Italian 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 2009/2010, 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 2010/2011 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 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 face-to-face training and some on-line sessions) 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. 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.⁶ Activities can be either remedial, thus targeting low achievers, or targeted to top-performing students. In our data, the average number of students involved as a proportion of class size varies between 25% and 100%; in more than 75% of classes at least 50% of the students participate in the afternoon activities. In most classes (about 65%) the number of activities chosen is between 2 and 4. This corresponds, on average, to 30–60 additional hours spent at school by participating classes over the school year. Overall, the choice of number of activities, kind of activities and proportion of participating students in each activity was done so to guarantee that each student within one class received the same amount of extra instruction time. Qualitative investigations on the implementation of the program (Mori 2011) show that, despite the general architecture of PQM, the attendance to training sessions was below the expectations and that Improvement Plans showed on average little to no relationship with the needs of the students highlighted in the standardized test that treated classes took in October. On the other hand, the organization of afternoon activities proceeded with no problems, so that we hypothesize that the main (or the only) channel through which the program may exercise an effect is the amount of additional instruction time.

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 INVALSI, which is in charge of testing Italian 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 2007/2008, and since the academic year 2009/2010 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. Our primary outcome is the result in the test score in mathematics and language. We consider also attitudes towards mathematics and language, study habits, motivation, anxiety during tests and school satisfaction. Appendix A (see online supplemental file at http://dx.doi.org/10.1080/09645292. 2015.1122742) provides a detailed description of the methods used to identify and create those measures, which we briefly describe in the following paragraph.

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 four items for mathematics and four items for language, which we summarize in one factor capturing attitudes towards mathematics and one factor capturing attitudes towards Italian language. Study habits are measured via 15 guestions 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'). Motivation is measured by seven items, which we summarize into two main factors through a factor analysis. The first measures 'External motivation', which captures a tendency to study and perform well in order to please parents or teachers, and the other one measures 'Internal motivation', which captures a tendency to study and perform well for personal satisfaction. School satisfaction is measured through a factor coming from five items battery asking about level of satisfaction with the school based on a 4-point scale. Finally, the guestionnaire contains a battery related to test anxiety, measuring agreement (4-point scale, as before) with seven items, which we summarize 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 2010/2011. This choice is driven by data availability: 2009/2010 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 2009/2010 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 analyses schools which participated in the programme in both years, thus concentrating just on schools which were selected for the first time in 2010/2011. 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 B-1 in the appendix (see online supplemental file at http://dx.doi.org/10.1080/09645292. 2015.1122742) shows descriptive statistics for all the schools in the four regions. The number of PQM schools varies among regions,⁷ with 10 schools in Calabria and around 40 in the other three. The table reveals a differential process of selection of schools in each region, but on average, the two groups of schools are not dissimilar over a wide range of variables. In particular, it is worth noticing that what seems to be common among the four regions is that none of the criteria used to choose PQM schools (retention and dropout rates) differentiate between the selected and nonselected schools.

We choose a group of control schools sharing similar observable characteristics to the schools enrolled in the programme through propensity score matching. Matching is used only as a preliminary tool to identify a set of schools (out of the universe of Italian lower secondary schools) that share similar observable characteristics with the treated ones.⁸

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.⁹ The variables used to estimate the propensity score are: average percentages of correct answers in mathematics and language in sixth grade in 2009/2010, student-teacher ratio, proportion of permanent teachers, dropout rate, fail rate, proportion of repeating students, proportion of students attending more than 30 hours per week, total number of students, use of PON funding in 2009/2010, population of the town and whether the school is located in a mountain municipality.¹⁰

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 1 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. This is confirmed also by some useful statistics, measured both before and after the matching, reported in the bottom part of the table. We see that after the matching the mean and median bias are lower than in the raw sample, the covariates are not jointly significant and the Pseudo R^2 has decreased. As a final check we plot presents descriptive statistics of average student characteristics for the pre-programme cohort among the two groups of schools (Table B-3 in the online appendix). The table shows that there are just minor differences between the two groups in terms of average individual characteristics, so the matching – even if was not based on student's covariates – managed to make the two sample comparable also in terms of students' characteristics.

Table 2 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 2010/2011.¹¹

	(1)	(2)	(3)	(4)
	PQM	Control	Difference	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	0.007	(0.007)	(2.385)
% of foreign students	0.027	0.027	0.000	1.088
0% of students with disabilities	0.024	0.021	(0.003)	(5.101)
% of students with disabilities	0.054	0.051	0.003	2.052
% of repeating students	0.049	0.041	(0.003)	(8.009)
% of repeating students	0.040	0.041	(0.007	(4.105
Drop out rate	0.003	0.003	(0.000)	(4.102) _5 368
Diop out late	0.005	0.005	(0.000	(12.07)
Failure rate	0 049	0.046	0.003	-3665
	0.019	0.010	(0.006)	(3.912)
% of classes doing more than 30 hours	0.335	0.337	-0.001	-0.170
5			(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		
Sample	Pseudo R ²	$p>\chi^2$	Mean bias	Median bias
Raw	0.044	0.001	16.3	14.4
Matched	0.017	0.975	6.6	5.8

Table 1. Descriptive statistics (schools enrolled in PQM and matched control schools) – pre-programme data only and matching quality statistics.

Notes: 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 is reported in parentheses; column (4) reports the results from a logit regression for being a PQM school (standard error in parentheses) using only PQM and control schools. In the bottom part of the table we report some statistics – before and after the matching – relevant to assess the matching quality.

Table 2. 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	6,228	6,461
Used as controls in PQM schools	8,260	8,380
Used as controls in remaining schools	12,455	12,672

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

5. Methods

So far we have selected, out of the universe of lower secondary schools in Italy, a control group of schools sharing similar observable characteristics with those enrolled in PQM with a matching procedure. In order to take into account unobservable characteristics and to provide an unbiased estimate of the effect of the programme, our identification strategy uses a difference-in-differences approach.

Using school identifiers provided by the INVALSI, we were able to link data for the same school in 2009/2010 (pre-programme period) and 2010/2011 (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 2009/2010 and 2010/2011. 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 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 classes, principals or schools.¹⁴ We are not able to identify which students who participate in which of the afternoon activities,¹⁵ thus the effect estimated is that of belonging to a class enrolled in the PQM. Our basic specification considers the following equation:

$$y_{ijt}^{k} = \beta_0 + \beta_1 \text{Math}_{jt} T_{jt} + \beta_2 \text{Italian}_{jt} T_{jt} + \beta_3 C_{jt} T_{jt} + \beta_4 T_{jt} + \theta X_{ijt} + \gamma_i + \epsilon_{ijt}^{k},$$

where y_{ijt}^k is the percentage of correct answers in subject k (mathematics or language) for student iin *sezione j* in year t. T_{jt} is an indicator of observations in the post-intervention year. Italian_{jt} and Math_{jt} are dummies for being enrolled in *any* activity in Italian language and mathematics, respectively, while C_{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_{jt}^k is a random error. We estimate two separate regressions, one for test score in mathematics (k = mathematics) and one for test scores in Italian language (k = Italian). The coefficients β_1 and β_2 are difference-in-difference estimates and are the main effects of interest, while β_3 captures possible spill over effects of treated classes on non-treated classes in PQM schools. 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). All standard errors are clustered at the school level.

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 by gender.

6. Results

Table 3 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

Table 3.	Effect o	of PQM	on the	percentage of	correct	answers in	mathematics	and Ita	alian la	anguage.

	-					
	N	lathematics	Itali	an language		
	(1) Whole sample	(2) Females	(3) Males	(4) Whole sample	(5) Females	(6) Males
Any extra class in mathematics	0.024** (0.012)	0.027** (0.013)	0.023* (0.013)	0.010 (0.009)	0.015* (0.009)	0.010
Any extra class in language	-0.010 (0.010)	0.001 (0.012)	-0.019* (0.011)	0.002 (0.008)	0.005 (0.009)	-0.002
Control class in PQM schools	-0.001 (0.009)	-0.002 (0.010)	0.002 (0.009)	-0.007 (0.007)	-0.009 (0.008)	-0.003
Observations	54,456	26,315	27,783	54,456	26,315	27,783

Notes: Difference-in-differences estimates of the effect of the intervention on mathematics (columns (1)–(3)) and Italian language (columns (4)–(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.

ahead or behind compared to his age), maximum level of education of the parents, mother working status and dummies indicating missing data (individual-level variables only; the class-level variables show no missing).

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; Zimmer, Hamilton, and Christina 2010; Jensen 2013). The estimated increase in percentage of corrected answers (2.4 for the overall sample) corresponds to an increase of 0.12 sd. The average increase in instruction time in mathematics, considering the average number of activities and the average proportion of students participating in each activity is roughly 25% (1 hour per week extra, being the weekly number of hours done in mathematics set to 4). The magnitude of our findings is not far from previous evidences, that find that an yearly increase of 25% of instruction time leads to an increase in mathematics test scores of 0.05 sd (Lavy 2012), 0.07 sd (Jensen 2013) and 0.15 (Lavy 2010).

As for spill over 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 the in 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.

12 👄 E. C. MERONI and G. ABBIATI

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, and the effects are not significantly different by gender. Indeed the programme does not manage to close the gaps between girls and boys in either mathematics or Italian language.¹⁶

	(1)	(2)	(2)	(4)	(5)	(6)			
	$(1) \qquad (2) \qquad (3)$			(4) (5) (6)					
	Positi	ve attitude t mathematio	cowards cs	Positive a	ttitude towa language	rds Italian			
	Whole			Whole					
	sample	Females	Males	sample	Females	Males			
Any extra class in mathematics	0.040	0.080*	0.019	-0.029	-0.048	-0.001			
	(0.034)	(0.048)	(0.045)	(0.042)	(0.048)	(0.053)			
Any extra class in language	-0.033	0.039	-0.114***	0.133***	0.116**	0.156***			
	(0.033)	(0.047)	(0.043)	(0.039)	(0.046)	(0.051)			
Control class in PQM schools	0.001	0.021	-0.007	0.022	-0.002	0.031			
	(0.025)	(0.035)	(0.034)	(0.031)	(0.037)	(0.037)			
Observations	49,992	24,204	25,488	49,704	24,073	25,330			
	(7) Ext	(8) ernal motiva	(9) ation	(10) Inte	(11) ernal motivat	(12) tion	(13) (14) Anxiety		(15)
	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	46,854	23,793	22,775	46,854	22,775	23,793	50,241	24,415	25,533
	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	Link and	remember i	nformation	Repeating while studying				Satisfaction	
	Whole sample	Females	Males	Whole sample	Females	Males			
Any extra class in mathematics	0.032	0.028	0.047	0.023	0.014	0.038	0.037	0.025	0.049
	(0.039)	(0.053)	(0.050)	(0.033)	(0.043)	(0.044)	(0.045)	(0.056)	(0.052)
Any extra class in language	0.045	0.092*	-0.008	0.007	0.050	-0.035	0.024	0.004	0.048
	(0.037)	(0.052)	(0.046)	(0.035)	(0.048)	(0.043)	(0.047)	(0.052)	(0.056)
Control class in PQM schools	0.016	0.016	0.018	-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	44,640	21,699	22,669	50,890	24,644	25,947	49,534	24,060	25,186

Table 4. Effect of PQM on attitude and behaviour.

Notes: Difference-in-differences estimates of the effect of the intervention on students' attitudes and behaviours, on the whole sample and on the sub-samples of females and males. 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.

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 4 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 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). 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 towards 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 suggest that receiving extra school hours in language negatively affects both performance and attitude towards 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. A possible explanation could be that 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, Harald Freudenthaler, and Neubauer 2010; Cornwell, Mustard, and Van Parys 2013; Fortin, Oreopoulos, and Phipps 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 towards 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 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 could act as a complement to language study time. Again, this result is consonant with findings in the existing literature: girls are characterized by greater motivation and discipline and more risk aversion (Steinmayr and Spinath 2008; Borghans et al. 2009; Spinath, Harald Freudenthaler, and Neubauer 2010; Cornwell, Mustard, and Van Parys 2013; Fortin, Oreopoulos, and Phipps 2013), leading them not to compensate the increased effort in one subject at the expensed of another one.

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. A possible explanation could be that if receiving extra instruction time in mathematics helps girls gain confidence in that subject, where they are usually behind, it can also lead to spill over effects on Italian language: girls are more motivated and committed, and seeing the results they are obtaining in mathematics could push them to invest more and thus perform better in Italian language too.

To provide additional evidence supporting this interpretation we estimated the effect of participating in PQM on a variable measuring the time students spend on non-educational activities after school. In the questionnaire students are also asked how many hours they devote to activities such as watching TV (including VHS and DVDs) and playing video-games. Children could answer:

14 👄 E. C. MERONI and G. ABBIATI

Table 5. Robustness checks.

none, less than 1 hour, 1–2 hours, more than two hours. We made these two variables continuous assigning a 0 to the first answer (none); 0.5 to the second (less than 1 hour); 1.5 to the third (1–2 hours); and 3 to the forth (more than 2 hours). Afterwards we summed them up and obtained one variable ranging from 0 to 6 which we called 'Time spent on non-educational activities'. We then estimate Equation (1) using this variable as dependent: the idea is to assess whether participation into PQM could have affected the time children devote to these kind of activities outside school time. Result for males and female are reported in Table B-4 in the appendix (see online supplemental file at http://dx.doi.org/10.1080/09645292.2015.1122742). While we see that participating in PQM had no effect on the time dedicated to non-educational activities after school for girls, we notice that for boys the reverse is true: those who participated to the Language programme increased the time devoted to TV and/or video-games. The increase is marginal because it correspond to an net increase of roughly 10 minutes per day, but it supports our hypothesis that for boys spending

A: Regression without control variables Mathematics Italian language (1) (2) (3) (4) (5) (6) Whole sample Females Whole sample Females Males Males Any extra class in mathematics 0.025** 0.029** 0.023* 0.011 0.018* 0.010 (0.012)(0.013)(0.014) (0.009) (0.009)(0.011)Any extra class in language -0.006 0.003 -0.015 0.006 0.008 0.003 (0.011)(0.012)(0.011)(0.009)(0.009)(0.010)Control class in POM schools -0.008 -0.001 0.002 -0.004 0.000 -0.007(0.008) (0.008)(0.010)(0.009)(0.007)(0.008)Observations 54,456 26,315 27,783 54,456 26,315 27,783 B: Regression within PQM schools only Mathematics Italian language (1) (2)(3) (4)(5) (6) Whole sample Females Whole sample Females Males Males Any extra class in mathematics 0.025** 0.030** 0.022* 0.016* 0.022** 0.013 (0.011)(0.012)(0.012) (0.008) (0.008)(0.010)Any extra class in language -0.009 0.003 -0.021* 0.008 0.013 0.000 (0.010) (0.012)(0.012) (0.008)(0.009) (0.010) Observations 29,329 14,258 14,877 29,329 14,258 14,877 C: Regression using the universe of Italian lower secondary schools Mathematics Italian language (1) (2)(3) (4) (5) (6) Whole sample Females Males Whole sample Females Males Any extra class in mathematics 0.027** 0.028** 0.027** 0.014* 0.018** 0.013 (0.012)(0.008)(0.008)(0.009)(0.011)(0.011)Any extra class in language -0.006 0.003 -0.014 0.006 0.009 0.001 (0.009)(0.011)(0.010)(0.007)(0.007)(0.009)Control class in POM schools 0.004 0.000 0.008 -0.002-0.005 0.001 (0.007) (0.008) (0.007) (0.005) (0.006) (0.006) Observations 246,211 118,779 125,915 246,211 118,779 125,915

Notes: Difference-in-differences estimates of the effect of the intervention on mathematics (columns (1)–(3)) and Italian language (columns (4)–(6)) percentage of correct answers, on the whole sample of students (column (1) and (4)), on the sub-sample of females (column (2) and (5)) and on the sub-sample of males (column (3) and (6)). Panel A re-estimate Equation (1) excluding control variables (class size, number of regular school hours per week, gender, migration status, regularity status, maximum level of education of the parents, mother working status and dummies indicating missing data); panel B re-estimate the same equation using as control classes only non-treated classes within PQM school, excluding from the analysis non-PQM schools; panel C re-estimate the same equation using as control schools the universe of lower secondary schools in the four targeted regions, disregarding the outcomes of the matching procedure. 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.

more time at school comes at the cost of reducing the time spend on educational activities at home. We know that the relationship between extra-school hours and leisure time may not be linear or simply additive. It shows, however, different reactions to extra-school time activities of the students according to their gender.

7. Robustness checks

This section is devoted to the robustness checks in order to further support our findings. First, we estimate Equation (1) without including control variables; then we estimate the model only within PQM schools, excluding control schools and using as control classes only non-treated classes in PQM schools; finally, as a third exercise, we replicate the analysis using as controls all schools in

Table 6. Robustness checks on attitude towards mathematics and Italia

	A: Regression without control variables								
	Attitude to	wards Mather	matics	Attitude to	wards Italian la	nguage			
	(1) Whole sample (1)	(2) Females (2)	(3) Males (3)	(4) Whole sample (4)	(5) Females (5)	(6) Males (6)			
Any extra class in mathematics	0.045	0.081*	0.022	-0.031	-0.043	0.001			
Any extra class in language	-0.020 (0.033)	0.045	-0.098** (0.043)	0.143*** (0.039)	0.121**	0.172***			
Control class in PQM schools	0.003 (0.026)	0.021 (0.036)	-0.002 (0.035)	0.026 (0.032)	-0.002 (0.037)	0.034 (0.038)			
Observations	49992	24204	25488	49704	24073	25330			
	B: Regression within PQM schools only								
	Attitude to	wards Mather	matics	Attitude towards Italian language					
	(1) Whole sample	(2) Females	(3) Males	(4) Whole sample	(5) Females	(6) Males			
Any extra class in mathematics	0.035 (0.033)	0.057 (0.047)	0.024 (0.045)	-0.050 (0.041)	-0.053 (0.047)	-0.029 (0.054)			
Any extra class in language	-0.035 (0.032)	0.018 (0.045)	-0.108** (0.044)	0.116** (0.039)	0.114** (0.047)	0.134** (0.049)			
Observations	26806	13096	13558	26713	13039	13522			
	C: Regression using the universe of Italian lower secondary schools								
	Attitude to	wards Mather	matics	Attitude towards Italian language					
	(1) Whole sample	(2) Females	(3) Males	(4) Whole sample	(5) Females	(6) Males			
Any extra class in mathematics	0.063** (0.031)	0.106** (0.044)	0.039 (0.042)	-0.032 (0.039)	-0.035 (0.045)	-0.019 (0.050)			
Any extra class in language	-0.009 (0.030)	0.066 (0.042)	-0.094** (0.039)	0.131**** (0.036)	0.129** (0.043)	0.137** (0.048)			
Control class in PQM schools	0.030 (0.019)	0.052 ^{***} (0.026)	0.020 (0.027)	0.020 (0.025)	0.014 (0.030)	0.009 (0.030)			
Observations	226,978	109,756	115,909	225,465	109.089	115.071			

Notes: Difference-in-differences estimates of the effect of the intervention on positive attitude towards mathematics (columns (1)– (3)) and Italian language (columns –(6)), 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)). Panel A re-estimate Equation (1) excluding control variables (class size, number of regular school hours per week, gender, migration status, regularity status, maximum level of education of the parents, mother working status and dummies indicating missing data); panel B re-estimate the same equation using as control classes only non-treated classes within PQM school, excluding from the analysis non-PQM schools; panel C re-estimate the same equation using as control schools the universe of lower secondary schools in the four affected regions, disregarding the outcomes of the matching procedure. 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<0.05.

***[.] *p*<0.01.

the four selected region, without employing the matching procedure as described in Section 4.2. In Table 5 we report the results from the first three exercises. In panel A we present the results of the estimation which does not include control variables used in the main specification: 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 dummies indicating missing data. Results are very similar to the one presented in the main specification (Table 3): we find a positive effect of receiving extra classes in mathematics on mathematics test scores, for both females and males, and a positive effect of receiving extra classes in mathematics test scores, the negative effect of receiving extra classes in ltalian to the original ones. Nevertheless, the negative effect of receiving extra classes in Italian on mathematics test scores for boys is no longer found (Coefficient is negative and similar in magnitude, but loses statistical significance).

In panel B we report the estimates of the effect of PQM only within PQM schools, using as control group only classes within PQM schools that were not engaged in extra-school time during the afternoon. All the results are confirmed. In panel C we report the results from the third exercise: using as control all classes in all the schools located in the four selected regions. Again, all the results are confirmed , with the only exception of the negative effect of receiving extra classes in Italian on mathematics test scores observed for boys, which is no longer found (as in exercise 1, the coefficient is negative and similar in magnitude, but loses its statistical significance).

Summing up, the results are robust to different model specifications and to different sample selections: positive effects of participating to PQM mathematics on mathematics test scores, for both males and females and a positive effect of participating into PQM mathematics on Italian Language for females only. Nevertheless the cross subject effect found for males (negative effects on mathematics test scores of receiving extra hours in language) is not found in the first and third exercise, raising some concerns about the statistical significance of this finding.

Similarly we replicate the results for the effect of the intervention on the attitude towards the two subjects under these three different settings¹⁷. Results, reported in Table 6 confirm that receiving extra time in language has a positive effects for both males and females on the attitude towards that subject, and that has a negative effect on the attitude towards mathematics for boys only. The effect found on the positive impact of receiving extra time in mathematics on positive attitude towards mathematics for girls is confirmed in specification B and C, while lack of significance under specification A. The general pattern of results on the attitudes is then confirmed.

8. 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; Zimmer, Hamilton, and Christina 2010; Jensen 2013), it seems to be easier for an increase in instruction time to improve achievement in mathematics than in 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. Additional results build on the previous work of Lavy (2012) and explore how an increase in instruction time could also affect non-cognitive outcomes. In the whole sample, we found no negative behavioural unexpected effects on students' behaviours or attitudes: on the contrary, we find a positive effect of receiving additional extra-hours in language on the positive attitudes towards studying 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 seems 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 towards 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. Cross-subject effects are rarely studied (or maybe rarely reported), even if there is some evidence of such phenomenon: Greenleaf et al. (2011), for instance, report an interesting case of a science professional development program for teachers exerting an effect also on language competences. It is difficult indeed to explain such pattern of results if we lack a theoretical model or previous robust evidence to build in. In this paper we provide a tentative explanation of why boys and girls react differently to extra-instruction time according to the subject. We must acknowledge, however, that our interpretation suffer from some limitations: some of this results are at the border of significance and not robust to all model specifications.

This suggests however that 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 better target educational interventions.

Notes

- 1. This project is financed by the EU PON Istruzione 2007–2013 fund (A-2-FSE-2009-2).
- 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.
- 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* B; a class of seventh 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.
- 6. Teachers receive 50 euro per hour gross. Thus, considering their salary, planning 4 activities would make one month's salary.
- 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, 2009/2010.
- 8. As better explained in Section 5 our identification strategy relies on a difference-in-difference approach.
- Only 4 PQM schools out of 138 had to be dropped because of this. Propensity score plots are reported in Figure B-1 in Appendix B (see online supplemental file at http://dx.doi.org/10.1080/09645292.2015.1122742).
- 10. The results of the logistic regressions are shown in Table B-2.
- 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.
- 12. Decreto ministeriale number 37, March the 26th, 2009.
- 13. This identification strategy could be undermined in case teacher mobility across schools was driven by the willingness to participate to the program. In the case of PQM, though, this event can be excluded: teachers have to apply for a change of school for the subsequent year from mid-January to late February (in this case, February 2010); the call for PQM 2010–2011 was opened in June 2010, and the list of eligible school among the applicants was made public in September 2010. Importantly, moreover, the call was not anticipated by any announcement by the Ministry. The timing of these bureaucratic procedures excludes both that mobility towards the future PQM schools and that teachers in (future) PQM schools applied for a school change less than their colleagues in other schools, in order to participate to the program.
- 14. The assumption needed for our identification strategy is that, in the absence of PQM, scores in all classes would have presented parallel trends. As mentioned, scores for sixth graders are available at the national level starting from the year 2009/2010. This makes it impossible to test for the existence of pre-intervention trends in the outcomes of interest using only our working sample. Another paper dealing with the same intervention (Battistin and Meroni)

2013), use scores for fifth graders that are available at the national level from 2008/2009, adding one year of preintervention data, and show that the trend can be considered as parallel.

- 15. As mentioned before we only have information on how many activities were proposed to the class and on how many student on average participated to each activity.
- 16. Table available upon request.
- 17. These were the only significant results found, a shown in Table 4

Acknowledgments

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. Part of the work carried out at the JRC's Centre for Research on Education and Lifelong Learning (CRELL) in the framework of the project KNOW (Human Capital for Prosperity and Sustainable Growth in Europe). We thank the two anonymous referees for their helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Elena Claudia Meroni b http://orcid.org/0000-0003-4904-9238 Giovanni Abbiati http://orcid.org/0000-0002-0236-8787

References

- Aucejo E., and T. Romano. 2014. "Assessing the Effect of School Days and Absences on Test Score Performance." Technical report, Centre for Economic Performance, London School of Economics and Political Science.
- 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. 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.
- Buser T., M. Niederle, and H. Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." The Quarterly Journal of Economics 129 (3): 1409–1447.
- Checkoway A., B. Boulay, B. Gamse, M. Caven, L. Fox, K. Kliorys, R. Luck, K. Maree, M. Velez, and M. Woodford. 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, and 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.
- Cortes K. E., J. S. Goodman, and T. Nomi. 2015. "Intensive Math Instruction and Educational Attainment Long-Run Impacts of Double-Dose Algebra." The Journal of Human Resources 50 (1): 108–158.

Croson R., and U. Gneezy. 2009. "Gender Differences in Preferences." Journal of Economic Literature 47 (2): 448-474.

- DeCicca P. 2007. "Does Full-Day Kindergarten Matter? Evidence from the First Two Years of Schooling." *Economics of Education Review* 26 (1): 67–82.
- Duckworth A. L., and M. E. P. Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." Journal of Educational Psychology 98 (1): 198–208
- 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.

Falch T., O. H. Nyhus, and B. Strøm. 2014. "Causal Effects of Mathematics." Labour Economics 31: 174-187.

- Fitzpatrick M. D., 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. H. H., 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* 41: 176–187.
- Goodman J. 2014. "Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time." Technical report, National Bureau of Economic Research.
- Greenleaf C. L., C. Litman, T. L. Hanson, R. Rosen, C. K. Boscardin, J. Herman, S. A. Schneider, S. Madden, and B. Jones. 2011. "Integrating Literacy and Science in Biology Teaching and Learning Impacts of Reading Apprenticeship Professional Development." American Educational Research Journal 48 (3): 647–717.
- 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. A., 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 Children's 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.
- Joensen J. S., and H. S. Nielsen. 2013. "Math and Gender: Is Math a Route to a High-Powered Career?." IZA DP No. 7164.
- 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 School's 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.
- Lee J.-W., and R. J. Barro. 2001. "Schooling Quality in a Cross-Section of Countries." Economica 68 (272): 465-488.
- 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. E. 2007. "Schooling and Test Scores: A Mother-Natural Experiment." *Economics of Education Review* 26 (5): 629–640.
- Marcotte D. E., and S. W. Hemelt. 2008. "Unscheduled School Closings and Student Performance." Education Finance and Policy 3 (3): 316–338.
- Martin A. J. 2004. "School Motivation of Boys and Girls: Differences of Degree, Differences of Kind, or Both?." Australian Journal of Psychology 56 (3): 133–146.
- 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.
- Mori S. 2011. "Analisi e progettazione di piani di miglioramento: uno studio basato sui dati pqm- piano nazionale qualitá e merito." Unpublished doctoral dissertation.
- Niederle M., and A. H. Yestrumskas. 2008. "Gender Differences in Seeking Challenges: The Role of Institutions." Technical report, National Bureau of Economic Research.
- Parinduri R. A. 2014. "Do Children Spend Too Much Time in Schools? Evidence from a Longer School Year in Indonesia." Economics of Education Review 41: 89–104.
- 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. New Brunswick, NJ: NIEER.
- Sims D. P. 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.

20 👄 E. C. MERONI and G. ABBIATI

- 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.
- Wagner P., B. Schober, and C. Spiel. 2008. "Time Students Spend Working at Home for School." *Learning and Instruction* 18 (4): 309–320.
- Wößmann L. 2003. "Schooling Resources, Educational Institutions and Student Performance: The International Evidence." Oxford Bulletin of Economics and Statistics 65 (2): 117–170.
- Xu J. 2006. "Gender and Homework Management Reported by High School Students." *Educational Psychology* 26 (1): 73–91.
- 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.