Teacher training, extra education and students' achievement. The evaluation of a program for promoting students' performances in Italian lower secondary schools

Elena Claudia Meroni
Department of Statistical Sciences
Padua University

Giovanni Abbiati
Department of Welfare and Labour studies
University of Milan

July 2012

Abstract

This paper evaluates the short-term effects of a program targeting students enrolled in sixth grade classes in lower secondary schools. The program provided a in-service training to the mathematics and language teachers of the treated classes and additional hours of education to the students belonging to these classes. The intervention aimed at rising students' performances in some of the worse performing schools located in four Italian southern regions. Using a comparison group of schools, chosen by propensity score matching, and implementing a difference-in-differences estimation strategy, we find that students in the treated classes reach higher test scores in mathematics, while no effect is found on the language test.

1 Introduction

The aim of this paper is to provide empirical evidence on the effectiveness of a program called PQM PON (italian acronym for National Plan for Quality and Merit)[1] on students' achievement. The program is designed to raise students' competences both by providing lower secondary schools teachers of language (italian) and mathematics with a training on the job and by organizing remedial and extra classes outside the normal scheduled activities for students. Hence, the program act on two relevant input dimensions: the

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

quantity of schooling received by the pupils, resulting in a higher numbers of hours and, on the other hand, the quality of it, being the teachers involved in a content-focused training providing innovative teaching materials and aming at rising teachers' didactic skills. The program was addressed to the more disadvantaged[2] lower secondary schools of the regions included in the PON funding (Campania, Siciliy, Calabria and Apulia) in the academic year 2010-2011.

Italy is an interesting case of study since national and international competence surveys (above all PISA 2006) revealed the existence of a serious gap between Italy and the rest of the OECD countries, and of a very wide gap between different geographical areas within the country. Students in the northern part perform in line with the rest of the European countries, while students in the South perform fairly below the averages. The South represents the poorest part of the country on a large set of economic and social indicators. This situation of socio-economic disadvantage is translated into the school system: international competence surveys have long since detected a large gap between the southern regions and the rest of the country, a gap that cumulates over time (as detected in the surveys IEA-PIRLS 2006; IEA TIMMS 2007; PISA 2003, 2006 and 2009). These evidences raised the interest towards the search for solutions for reducing the gap within the different areas of the countries, and hence within the OCDE countries. The paper is structured as follow: section 2 provide a review of the literature dealing with the effect of both teacher professional training and remedial and extra education on students' achievement; section 3 provide a detailed explanation of the PQM program and a brief explanation of the Italian school system; section 4 explains the methodology used to find the causal effect of the PQM program on students' achievements; section 5 explain the data used and provides some descriptives statistics; section 6 presents the results and finally section 7 concludes.

2 Related literature

Since the recognition of the variability in student achievement due to teachers, a lot of research has been devoted to the identification of the core features of an effective teacher and to how to improve them. There is large consensus about the centrality of teachers in the achievement process, but, as Goldhaber (2002) has pointed out, only a residual part of the variation of the teacher effect on students' achievement can be attributable to observable characteristics. If studying the effect of teacher characteristics on students outcomes raise too severe estimation issues, it is possible, however, to study the effectiveness of particular training courses, that can act on the non-innate component of teaching ability. Professional development, either to

[2]Schools with higher rate of drop out, repeating and failing students

be done through pre-service or in-service training, is the key mechanism for the development of both didactic skills and content knowledge. Teachers who possess qualification certificates are found to be more effective than teachers without any certificated qualification (Goldhaber, 2007; Darling-Hammond, 2000; Garet et al., 2010).

Recently, empirical evidence has been gathered also on larger scale randomized control trials and extending research on middle and secondary schools. Even if these programs have been built according to the features of successful professional development identified by Garet et al. (2001), the results are quite mixed. While two programs involving science teachers, both at primary (Thurston et al., 2008) and middle schools (Johnson et al., 2007), found positive effects of the training on students' outcomes, others did not. Santagata et al. (2010) found no effect of a specific training program on middles school math and science teachers. The research team coordinated by Garet set a large scale randomized experiment on middle school math teachers providing instruction on pedagogy, content knowledge, and resource materials, during both a summer school sessions and school-year seminars. Results after the first and second year of implementation showed no significant impact of this in-depth training on students achievement (Garet et al., 2001, 2010). The same absence of significant or substantial effects was found also on another large scale experiment on primary school language teachers (Garet et al., 2008).

As for research dealing with remedial and extra education, the focus has mainly been on the effect of remedial classes on both students grade and students probability to enroll into higher education, hence focusing on high school or college students. Again findings are quite mixed. Aiken et al. (1998) were among the firsts in trying to assess the causal effect of remediation: they could apply three designs (randomized experiment, nonequivalent control group design, regression discontinuity design) to study the effect of a university-level freshman remedial writing program. The authors find a positive effect of the program, in spite of the small size of the sample. Bettinger and Long (2009) using an instrumental variable approach are able to identify a positive causal effect of remediation on college outcomes (performances and persistence) of under-prepared students in Ohio public colleges. On the other hand, Calcagno and Long (2008) using a regression discontinuity approach on a sample of more than 100,000 students in Florida, find that the remedial courses increase the probability of completing the first year of college, but they do not affect the likelihood of completing the whole degree. Positive results of remedial high school programs were found also on courses addressed to younger high school students (Lang et al., 2009) and outside the US context (Lavy and Schlosser, 2004). For what concerns summer schools programs, Jacob and Lefgren (2004) perform a regression discontinuity design to identify the effect on academic achievement of a program addressed to younger students, finding a positive effect on third graders but no effect on sixth graders. A previous randomized study study on a summer school program including summer employment and academic remediation aimed at contrasting early school dropouts, instead, found no effects (Grossman et al., 1992). It is worth nothing that much of the evidence presented focuses on remedial education provided during regular time only. As Lavy and Schlosser (2004) point out, it is a bit surprising that courses held outside regular time received so few attention.

The effect of remedial classes in Italy has been recently studied on a sample of students in upper secondary schools (Battistin et al., 2010). Using a quasi experimental variation resulting from a geographical discontinuity in the implementation of the remedial education reform introduced in upper secondary schools in Italy (the so-called Fioroni reform, implemented from the academic year 2007-2008), the authors find a positive effect of remedial classes on the performances of the academic track students, while it has a negative effect on students attending vocational high schools.

This review of the robust findings on remedial education show how empirical evidence is scarce, almost totally focused on high school students. Even if the majority of them finds an effect either on student achievement or on grades, the heterogeneity of the phenomenon put severe issues in generalizing the results from single-program studies. The studies, however, are promising in light of the findings that suggest a bigger impact of programs targeted to students at early life course stages compared to high school (Heckman and Carneiro, 2003). Middle and primary school remedial programs, although seldomly implemented, could thus represent a potentially effective channel to raise student achievement.

3 The PQM program in the Italian school system context

3.1 The Italian school system

The italian school system is structured as follows: students attend the primary school from grade first to fifth, then from grade sixth to eight they attend the lower secondary school. The programs taught in primary and lower secondary schools are the same for all the students are settled by the Italian Ministry of Education. At the end of the eight grade students start the higher secondary school and are free to choose among three main different major tracks: the vocational high school (Istituto professionale and Corsi di formazione professionale), the technical high school (Istituto tecnico) and the academic one (Liceo).

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

In theory assignment, of both teachers and students, to the different *sezioni* shuold be random, but in practice it is well known that there are some mechanisms (parents' pressure to have their children in a given *sezione*, school principal assigned some teachers to a given *sezione*, ...), which could lead to a different composition of the different *sezioni* inside a school.

For what concern teachers professional development, the italian system has some peculiar characteristics. Unlike most OECD countries, in fact, Italy only recently has organized teacher training to transmit both disciplinary and pedagogical skills to future teachers. Until 1998, to become a primary school teacher was requested just a 5 years upper secondary school diploma. For secondary school teachers was instead required a university degree. The access to teaching took place through a disciplinary exam where no attention was given to the pedagogical content of the profession. After 1998, future primary schools teachers had to graduate at the Primary Education Sciences faculties. In order to become a secondary school teacher, university graduates had to follow two years of postgraduate courses in a SSIS (italian acronym for Secondary School Teaching Post-Graduate Course)[3].

Despite the fact that in last 14 years every aspirant teacher received some basic training also on pedagogical and didactical skills, the little number of new-enrolled teachers made permanent did not manage to change the general picture of Italian teachers. The vast majority of them received no training at all before being hired, and, during the first years of work, they had to rely only on the help of tenured colleagues, their short experience and their own personal teaching ability. National and international surveys on teachers (Cavalli, 1992, 2000; Cavalli and Argentin, 2010) show a growing demand of professional development of italian teachers, who are conscious of the limits of their pre-service training, especially for what concerns didactic and ICT skills. Along with an increasing demand, teachers express also a general dissatisfaction

[3] In 2010 the SSIS has closed, replaced for the school year 2010-2011 by a year of apprenticeship. In 2011 the apprenticeship year have been suppressed. At the time we are writing it is unclear the future organization of the apprenticeship/postgraduate courses for aspirant secondary school teachers.

towards in-service training, that fail in the aim to be a sufficient plugging device towards the lacks of previous training.

3.2 The regional differences between the North and the South

In the context of research in education, the South of Italy is an interesting field of study for various reasons. The South represents the poorest part of the country on a large set of economic and social indicators. This situation of socio-economic disadvantage is translated into the school system: international competence surveys have long since detected a large gap between the southern regions and the rest of the country (IEA-PIRLS 2006; IEA TIMMS 2007; PISA 2003, 2006 and 2009). Moreover recent national tests show that while the North and South differences among second graders are of minor entity, they increase at the end of the primary school and they grow again during the middle school (INVALSI, 2010a).

A closer analysis of the results reveals that much of the disadvantage of the South can be attributable to the big differences at the school level: the variance between schools account for the 26% of the total variance among sixth graders, against a national mean of 19% and a mean of 9% in the northern regions (ibidem). In secondary schools, a significant part of the variance between schools seem to be attributable to the socio-economic status of the students (Alivernini et al., 2010). The school context seem thus to play a crucial role in shaping the different performances of the students (Bratti et al., 2007).

For these reasons four of the Italian regions located in the South (Campania, Siciliy, Calabria and Apulia) received EU funding as part of the Programma Operativo Nazionale (PON) to improve teaching and learning processes in middle and high schools. One of the actions taken with these fundings was the implementation of the PQM program.

3.3 The PQM Program

PQM PON is set up as an integrated plan of activities addressed to lower secondary schools of the regions included in the PON funding. The program involves both training on the job for mathematics and language (italian) teachers and remedial and extra classes for students, to be held outside regular school time. The main objective of PQM is to increase the students' performances in the two subjects.The target population of the PQM program are sixth grade students, in classes taught by trained teachers, in the selected schools. The project, set up during the spring/summer of 2009, was implemented in the school year 2009-2010 in 304 schools only for mathematics teachers. During the following school year the program was extended also to italian teachers and more schools were involved. In this paper we will focus only on the schools starting the project in the second year, given the the lack of availability of test scores measures of sixth graders in italian and mathematics pre and post treatment for the first year of implementation. Moreover we exclude schools who participate in the program both in 2009-2010 and in 2010-2011, since our evaluation strategy rely on a difference-in-difference estimator, and therefore we couldn't keep the sample schools who were exposed to the treatment for two consecutive years.

The selection of the schools involved in PQM is based on a ranking of schools that decided to apply to the program. Each school could apply and then in each region, the Regional School Offices had to decide which schools would actually participate. The ranking was built using the following indicators: percentage of dropped-out students; percentage of failing students; percentage of repeating students. The selection process intended to favor the more disadvantaged schools among the applicants and the majority of the schools applied both for math and language program in all the four region considered.

The program is structured as follows: once a school is selected into PQM, two teachers per disciplinary area (mathematics or italian) are chosen to follow the on the job training (30 hours of formal training and 30 hours of online training to his class). The training has two main goals. First, to provide teachers with innovative teaching material, designed to give alternative solution to teach the usual curricular contents using elements such as games, team-work and lab activities. Second, to help the teachers to set up a *Improvement Plan*, based on the return of the results of the standardized test which treated classes take at the beginning of the academic year (October). Once the plan is settled, the teachers start experimenting the planned activities in the PQM classes, both during regular school time and during the extra classes in the afternoon. The afternoon activities planned per school range from a minimum of three to a maximum of seven[4]. Each activity foresees an average of 15 hours of remedial and extra education to be held outside the regular school time to small groups of students. In May, students are tested a second time with the standardize test taken by all the sixth graders in the country. The program is thus not intended to be a traditional content-focused training program but a polyvalent training aimed at providing teachers with diagnosis instruments, didactic planning skills and didactic materials.

4 Methods

The key interest of the paper lies in estimating the causal effect of the treatment on student's test scores after one year. The unit of interest are treated classes, that is classes in the school selected to participate in

[4]Differently from normal training programs, teachers receive extra-salary according to the number of remedial activities.

the program chosen to be taught by a trained teacher and to follow the extra hours in the afternoon. Given the definition of treatment we will have two kinds of units which are not exposed to the treatment: classes in pqm schools which are not selected to be PQM classes, and classes in schools which are not selected to be PQM schools.

The first issue we need to deal with is the non random selection of schools into the PQM program, indeed as explained in section 3, the program is directed to lower performing schools. In order to partly eliminate the bias resulting from the non random selection of the school into the program, we choose a group of control schools, among all the non pqm schools, which share similar observable characteristic with the pqm schools. In practice we match pqm schools with non pqm schools located in the same province, using propensity score matching with replacement, calculated as the probability of being assigned to the pqm program conditional on a set of observable variables measured at the school level, before the program started (year 2009). Variables included are: average test scores in mathematics and language; student-teacher ration, proportion of permanent teachers, drop out rate, failing rate, proportion of repeating students, proportion of immigrant students, proportion of disable students, proportion of female students, proportion of students attending more than 30 hours per week, average class size, number of students, whether the school has received in the previous year other PON funds for other activities and population in town.

Once we have matched the schools, we estimate a standard difference-in-differences model, in which we observe average performances in the treated and non treated classes (both in pqm schools and in control schools), before and after the implementation of the program. Notice that we do not follow the same children over time (i.e. with do not have panel data at the student level), but we compare performances of two contiguous cohorts of children, belonging to the same *sezione* at the end of sixth grade, before and after the program implementation. The assumption needed to identify the causal effect of being in a treated classes are two: the first is the usual common trend assumption, that is treated and non treated classes follow the same time trend; the second is that we allow teachers and students to be assigned non randomly to a given class(*sezione*) inside a school, but we assume that this non random assignment is constant trough time. For example, assume that in a school there are 3 sixth grade classes (3 *sezioni*) in year 2009-2010, class A, B and C; what we assume is that if the best teachers (students) are assigned to class (*sezione*) A in 2009-2010, they will be assigned to class (*sezione*) A also in the following year. This is a very plausible assumption in the Italian schools context, and it can be shown that indeed there exist a rank correlation between classes (*sezioni*) through time.

The estimated model is than the following:

$$y_{jst} = \alpha_0 + \alpha_1 T_t + \alpha_2 D_{js} + \alpha_3 T_t * D_{js} + \alpha_4 T_t * S_s + \beta X_{jst} + \delta_s + \epsilon_{jst},$$

where y_{jst} is the average test score in mathematics or language in class (*sezione*)j, in school s a the end of year t; T_t is the dummy for time, where $T_t = 1$ stands for post-program period (year 2010-2011); D_{js}, is a dummy for treated classes, which takes value 1 in treated classes; the coefficient associated to the interaction of these two terms, $T_t * D_{js}$ is the parameter of interest, that is the difference-in-difference estimator, capturing the effect of being in a treated classes in the post program period; S_s is dummy which is equal to 1 for all the classes in pqm schools, both treated and non treated, and it appears in the interaction with T_t, this interaction capture eventual spill over effects of the program in non treated classes in the pqm schools. Finally, X_{jst} is a vector of average class characteristic, including proportion of female, of immigrants, of ahead and repeating students, average socio economic status, class size, and an indicator for whether the class is doing more than 30 hours; δ_s is a school fixed effect, and ϵ_{jst} is the error term.

5 Data

Data at the school level are provided by the Italian Ministry of Education, through INVALSI. We have access to a very detailed collection of databases containing information about all the lower secondary schools in Italy. In particular we have a database containing general information about the schools, from which we extract variables that we use in the calculation of the propensity score matching. Moreover for all the schools we know the exact municipality in which they are located, hence we can use this information to control for general geographical and demographic characteristics of the environment where the schools operate.

Data at the student level are collected directly by the INVALSI, which is in charge of testing the Italian students' performances trough a series of standardized tests in mathematics and language. These tests were introduced on a small sample of schools in the second and fifth grade in the academic year 2007-08, and since the academic year 2009-2010 these tests are taken by all students in Italy at the end of second, fifth, sixth, eight grade. This survey contain information on the results of the standardized tests, both for mathematics and language, the main socio-demographic characteristics of the child and his family (gender, year of born, origin, level of education and employment status of the parents, household composition, socio economics status) and questions about motivation and perception of the school and the way of studying of each student.

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

Details about the schools who are involved in PQM are provided by ANSAS, thanks to which we can identify participating schools and targeted classes.

The outcome variables used for this paper are standardized test score of sixth graders in mathematics and italian collected at the end of the school years 2009-2010 and 2010-2011, the former being the pre-treatment and the latter the post-treatment outcomes. Language test are built to measure reading proficiency (in particular the ability of the students to understand and interpret a text) and lexical and grammatical knowledge, while mathematics text is measuring knowledge of the mathematics contents and logical and cognitive processes used in the mathematical reasoning.

The tests are composed mainly by multiple choices questions, in which the students have to select the right answer out of four possibilities; in mathematics tests there are also few open questions. The test score is calculated simply as percentage of corrected answers out the total number of questions (42, in 2010 and 48, in 2011 for mathematics and 58, in 2010 and 63, in 2011 for language), and hence varies between 0 and 100. In the main equation, test scores have been standardize, in each year, to have mean 0 and standard deviation 1.

In Tables 1, 2, 3 and 4 we find descriptive statistics for all the schools in the four regions. The number of pqm schools varies between the four regions [5], with just 11 schools in Calabria and around 40 in the other three regions. The tables reveal a differential process of selection of the school inside each region. In Campania (table 1), the two group of schools are not dissimilar on wide range of geographical as well as school level variables, except for the student-teacher ratio and the school size: PQM schools are bigger and with a higher student-teacher ratio. Apulia and Calabria show two opposite situations: while in Apulia PQM schools perform worst than non-PQM schools, in Calabria they were selected among the better schools. Students of PQM schools in Apulia, in fact, score on average 2 percentage points less in both the 8th grade test 2009-2010 and in the 6th grade test 2009-2010, for language and math. PQM schools, moreover, show a higher rate of failing and repeating students than non PQM schools. The reverse situation characterize Calabria schools, where PQM schools score fairly higher than non PQM schools. In addition the index of Socioeconomic status indicates a substantive advantage for PQM schools. Finally, in Sicily we observe a mixed situation: PQM schools score on average higher in language but lower in math. What seem to be

[5] These numbers take into account that we dropped all the schools who were doing the pqm program also in the pre-treatment year , 2009-2010

common among the four region is school size: PQM schools are bigger than non PQM schools, probably because of the requirements (they had to be at least two tenured teachers).

In table 5 we report the estimates of the four logistic regressions made in the four regions to calculate the propensity scores: in each region, we estimate the probability of being selected for the PQM program, conditional on the observable characteristics at the school level. As previously seen, the patterns of association vary in every region: students results at national tests are loose predictors of the treatment, as well as many of school context variables (student teacher ratio, class size. proportion of females and tenured teachers). As previously seen, the number of students of a school seem to play a role, but it significant only in Sicily. Variables such as Socioeconomic status and proportion of students staying 30 hours at school are negatively correlated with the PQM status, especially in Calabria. Some variables at the school level show a strong association with the treatment, even if not statistically significant mainly for the low numbers involved (especially in Calabria): rate of failing and repeating students, school drop-out rate and proportion of disabled students. Even for these factors, however, the strength and the direction of the coefficients vary strongly from one region to another. This heterogeneity supports our choice to match PQM schools and controls separately by region: this way we manage not only to drop residual contextual effects that we do not capture via our model, but also it delete a potential unobservable selection processes at regional level.

Once obtained the propensity score, we matched each PQM school with the non treated school located inside the same province with the closer propensity score [6].

In table 6 and 7 we find the final numbers of schools, classes and students in PQM and schools chosen as control, in both pre and post treatment year. Finally in table 8 we report the same descriptive statistics for the PQM schools and for the schools chosen as control. There are a total of 134 PQM schools and 115 schools chosen as control. The mean of the variables in the two groups of schools (columns 1 and 2) are very similar; in column 3 we report the estimates of a logistic regression for the probability of being a PQM school, using as observation just PQM schools and schools chosen as control. After the matching none of the variables has any effect of the probability of being a PQM school, which is a suggestion of the fact that the matching was successful in choosing a group of schools with similar observable characteristics.

[6] Matching was done with replacement, hence more PQM schools could be matched with the same control school

6 Results

In table 9 we report the estimates for the model. What we observe is that the program seems to have a positive and significant effect on mathematics test score, while no effect is present for the language test score. We find a positive effect of being in a class assigned to the program on the italian test score. This, however, could simply capture the fact that for example, classes chosen to participate in the PQM language program were already the best classes in schools, thing which is not true for mathematics classes. We find no evidence of spillover effects, since the interaction between being a PQM school and being in the post-treatment period is not significant.

As for the other variables, we observe that a higher proportion of female in the class is correlated with an higher language test scores, while correlated with a lower mathematics test score. This result is in line with other findings about gender differences in performances in different subjects, which indicate boys as having an advantage in mathematics and a disadvantage in reading. The other variables behave in line with expectation, since a higher proportion of foreign students and behind students is negatively correlated with both subject, and the size of the coefficient is a bit higher on the language test score, while an higher proportion of ahead students show a positive association with the score in both subjects; a higher socio-economic status is associated to better performances in both subjects; doing more than 30 hours per week has a positive effect on mathematics and being in a larger class have a positive effect on both mathematics and language.

In order to provide further evidence of the real effect of the PQM program on students' performances we checked whether the program had any effect on the other inputs that enter the educational production function. For example if we find that the difference-in-difference estimator has an effect in reducing the number of behind students in treated classes, or increasing the proportion of male, it may be that the effect that we are finding is simply the effect of other inputs which have change consequently to the implementation of the program, violating the common trend assumption. Therefore we estimate different regressions using as dependent variables the main inputs that we use as controls in the main regression: proportion of females, proportion of immigrants, proportion of ahead students, proportion of behind students, average socio-economic-status, class size and whether the class is doing more that 30 hours per week (tables available by request to the authors). We find that the in none of the cases the interaction between being in a treated class in the post period program has any effect on the inputs. Therefore the program has no effect in the other variables that we use as control in the main equation.

7 Conclusion

This paper tries to establish the effectiveness of a program implemented in some lower secondary schools in four southern Italian regions. The targeted schools were the lower performing ones, and the program act on two dimensions since it both provide teachers with a training of the job and students with extra hours of education in the afternoon.

Using a propensity score matching combined with a difference-in-difference strategy with find a positive effect of the program on mathematics test score, while no effect is found on language test score. Nevertheless this results it is not surprising given the structure of the program which included extra time at school and remedial classes, and gave the teacher new methodologies to use in the classes. Indeed it is expectable that these components will affect more, at least in the short run, a subject which involves a lot of exercise and practice, like mathematics, rather that a more complex subject which mainly involve reading skills and compression, like language, things which are more hard to influence and shape simply by adding more hours of school.

8 Tables

Table 1: Descriptive at the school level, Campania

	Non pqm schools	Pqm schools
Italian, average test score, sixth grade	58.21	57.83
Mathematics, average test score, sixth grade	48.66	48.87
Italian, average test score, eight grade	53.78	54.61
Mathematics, average test score, eight grade	45.57	45.49
Proportion of permanent teachers	0.890	0.917
Student-teacher ratio	9.032	9.933
Number of students in the school	334.2	430.3
Proportion of immigrant students	0.0246	0.0222
Proportion of disable students	0.0338	0.0307
School drop out rate	0.00180	0.00141
School rate of failing students	0.0399	0.0405
School rate of repeating students	0.0365	0.0382
Average socio economic status	0.0173	-0.101
School received PON funds for students'activities	0.859	0.930
Proportion of female in the school	0.479	0.492
Class size	20.86	21.67
Proportion of students doing more than 30 hours per week	0.385	0.394
(Log) population in town	9.956	10.12
Number of schools	461	43

Table 2: Descriptive at the school level, Puglia

	Non pqm schools	Pqm schools
Italian, average test score, sixth grade	59.73	57.82
Mathematics, average test score, sixth grade	50.25	47.78
Italian, average test score, eight grade	57.65	55.19
Mathematics, average test score, eight grade	50.85	48.63
Proportion of permanent teachers	0.901	0.887
Student-teacher ratio	10.31	10.34
Number of students in the school	349.5	438.9
Proportion of immigrant students	0.0266	0.0279
Proportion of disable students	0.0262	0.0317
School drop out rate	0.00182	0.00256
School rate of failing students	0.0327	0.0385
School rate of repeating students	0.0323	0.0403
Average socio economic status	0.246	0.309
School received PON funds for students'activities	0.960	1
Proportion of female in the school	0.484	0.480
Class size	22.28	22.49
Proportion of students doing more than 30 hours per week	0.314	0.262
(Log) population in town	9.969	10.61
Number of schools	253	43

Table 3: Descriptive at the school level, Calabria

	Non pqm schools	Pqm schools
Italian, average test score, sixth grade	55.90	60.07
Mathematics, average test score, sixth grade	47.81	50.80
Italian, average test score, eight grade	52.04	56.98
Mathematics, average test score, eight grade	44.84	47.87
Proportion of permanent teachers	0.812	0.903
Student-teacher ratio	7.454	9.239
Number of students in the school	218.6	354.7
Proportion of immigrant students	0.0370	0.0363
Proportion of disable students	0.0296	0.0230
School drop out rate	0.00346	0.00466
School rate of failing students	0.0444	0.0341
School rate of repeating students	0.0375	0.0483
Average socio economic status	-0.0195	0.540
School received PON funds for students'activities	0.879	0.909
Proportion of female in the school	0.478	0.485
Class size	19.74	20.09
Proportion of students doing more than 30 hours per week	0.522	0.242
(Log) population in town	9.008	10.21
Number of schools	231	11

Table 4: Descriptive at the school level, Sicilia

	Non pqm schools	Pqm schools
Italian, average test score, sixth grade	55.41	54.57
Mathematics, average test score, sixth grade	46.74	45.72
Italian, average test score, eight grade	53.03	55.22
Mathematics, average test score, eight grade	45.22	43.41
Proportion of permanent teachers	0.842	0.855
Student-teacher ratio	8.556	8.655
Number of students in the school	297.1	356.3
Proportion of immigrant students	0.0337	0.0287
Proportion of disable students	0.0399	0.0438
School drop out rate	0.00398	0.00478
School rate of failing students	0.0646	0.0689
School rate of repeating students	0.0699	0.0637
Average socio economic status	-0.0657	-0.129
School received PON funds for students'activities	0.875	0.976
Proportion of female in the school	0.485	0.504
Class size	20.71	20.67
Proportion of students doing more than 30 hours per week	0.408	0.371
(Log) population in town	10.13	10.51
Number of schools	423	42

Table 5: Probability of being a PQM School

	Campania	Puglia	Calabria	Sicilia
Italian, average test score, sixth grade	-0.087	0.013	0.106	-0.011
	(0.054)	(0.064)	(0.128)	(0.049)
Mathematics, average test score, sixth grade	0.032	-0.076	0.008	-0.000
	(0.036)	(0.060)	(0.079)	(0.035)
Italian, average test score, eight grade	0.024	-0.012	0.080	0.059*
	(0.034)	(0.027)	(0.059)	(0.027)
Mathematics, average test score, eight grade	-0.026	-0.027	0.022	-0.038
	(0.027)	(0.033)	(0.058)	(0.021)
Proportion of permanent teachers	3.088	-3.559	2.636	0.066
	(2.344)	(2.077)	(4.066)	(1.701)
Student-teacher ratio	0.122	0.123	-0.390	-0.159
	(0.131)	(0.167)	(0.366)	(0.146)
Number of students in the school	0.002	0.002	0.004	0.003*
	(0.001)	(0.001)	(0.003)	(0.001)
Proportion of immigrant students	37.041	-0.770	14.079	0.221
	(24.227)	(14.568)	(13.799)	(3.804)
Proportion of disable students	-4.141	23.334	-39.918	5.530
	(11.852)	(14.664)	(32.327)	(6.777)
School drop out rate	26.677	3.952	14.476	8.448
	(55.833)	(36.572)	(28.119)	(12.683)
School rate of failing students	-0.660	-9.622	-31.166	3.627
	(5.540)	(8.354)	(16.385)	(3.923)
School rate of repeating students	1.187	6.416	34.247*	-7.130
	(5.958)	(8.787)	(14.219)	(4.322)
Average socio economic status	-0.140	0.805	1.742*	-0.079
	(0.203)	(0.476)	(0.789)	(0.226)
School received PON funds for students'activities	0.873		0.515	1.697
	(0.642)		(1.422)	(1.064)
Proportion of female in the school	3.343	-1.446	-2.884	3.743
	(2.692)	(3.428)	(5.563)	(2.547)
Class size	-0.020	-0.044	-0.187	-0.007
	(0.048)	(0.050)	(0.104)	(0.043)
Proportion of students doing more than 30 hours	0.579	-0.106	-2.625*	-0.236
	(0.479)	(0.677)	(1.326)	(0.511)
(Log) population in town	-0.123	0.219	0.516	0.035
	(0.139)	(0.166)	(0.378)	(0.131)
Any variable at the school level missing	-0.106	0.060		-0.059
	(1.097)	(0.967)		(0.793)
Constant	-4.956	2.926	-15.289	-6.482*
	(3.609)	(3.842)	(8.413)	(3.144)
Number of schools	504	286	232	465

The four columns correspond to four different regressions, one in each region, for the probability that a school is assigned to PQM. Estimates are at the school level, using pre-program characteristics

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

Table 6: PQM Schools

	Pre treatment year	Post treatment year
Number of schools	134	134
Number of treated classes	151	151
Number of treated students	3117	2942
Number of control classes	570	570
Number of control students	11949	11553

Table 7: Control Schools

	Pre treatment year	Post treatment year
Number of schools	115	115
Number of classes	581	581
Number of students	12083	12517

Table 8: Descriptives at the school level of pqm and control schools, after the matching

	PQM	Control	Probability of begin a pqm school
Italian, average test score, sixth grade	57.297	57.909	-0.022
			(0.042)
Italian, sd of test scores, sixth grade	14.627	14.437	0.055
			(0.084)
Mathematics, average test score, sixth grade	48.080	48.825	0.005
			(0.032)
Mathematics, sd of test scores, sixth grade	16.444	16.458	-0.010
			(0.077)
Italian, average test score, eight grade	55.111	54.849	0.015
			(0.022)
Mathematics, average test score, eight grade	46.083	46.451	-0.011
			(0.020)
Proportion of permanent teachers	0.891	0.891	0.201
			(1.520)
Student-teacher ratio	9.611	9.801	-0.101
			(0.108)
Number of students in the school	395.799	387.304	0.001
			(0.001)
Proportion of immigrant students	0.027	0.028	0.080
			(5.085)
Proportion of disable students	0.034	0.033	-3.860
			(7.768)
School drop out rate	0.003	0.002	13.348
			(17.290)
School rate of failing students	0.048	0.049	0.027
			(4.214)
School rate of repeating students	0.047	0.048	-1.165
			(4.495)
Average socio economic status	0.091	0.104	-0.010
			(0.188)
School received PON funds for students'activities	0.963	0.965	-0.013
			(0.713)
Proportion of female in the schools	0.494	0.491	1.520
			(2.224)
Proportion of classes doing more than 30 hours	0.337	0.333	-0.084
			(0.395)
Average class size	21.813	22.058	-0.018
			(0.056)
Campania	0.321	0.330	0.032
			(0.538)
Puglia	0.299	0.270	0.269
			(0.545)
Sicilia	0.299	0.313	-0.115
			(0.558)
Constant			0.463
			(3.005)
Number of pqm schools	134		
Number of control schools	115		

Descriptive statistics at the school level in PQM and schools chosen as controls, using pre-treatment year covariates.

In the third column estimates of the probability of being a PQM schools, estimated using just PQM and schools chosen as control

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

Table 9: Effect of pqm on the test scores

	Language	Mathematics
Treatement*post, mathematics		0.131*
		(0.057)
Treatement*post, language	0.016	
	(0.047)	
Post	0.057*	0.037
	(0.027)	(0.028)
Treatement, mathematics		-0.005
		(0.043)
Treatement, language	0.072*	
	(0.034)	
School pqm*post, mathematics		0.017
		(0.051)
School pqm*post, language	0.012	
	(0.041)	
Proportion of female in the class	0.161*	-0.255***
	(0.078)	(0.075)
Proportion of foreign in the class	-0.519**	-0.551*
	(0.174)	(0.235)
Proportion of ahead students in the class	0.836***	0.698***
	(0.186)	(0.194)
Proportion of behind students in the class	-1.091***	-0.723***
	(0.172)	(0.168)
Average socio economic status in the class	0.165***	0.166***
	(0.026)	(0.025)
Class size	0.019***	0.015***
	(0.003)	(0.003)
More than 30 hours per week	0.068	0.115**
	(0.037)	(0.043)
Constant	-0.499***	-0.242**
	(0.084)	(0.088)
Observations	2604	2604

Robust standard error clusterd at the school level in parenthesis

Estimates at the class level, with school fixed effects

Test scores have been standardized

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

References

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

F. Alivernini, B. Losito, and L Palmerio. *Le differenze nelle prestazioni degli studenti italiani in PISA 2006 e l'equit del sistema scolastico italiano : due studi.* Armando, 2010. Problemi dell'educazione, 2010.

E. Battistin, E. Covizzi, and A. Schizzerotto. The effects of remedial exams on student achievement: Evidence from upper secondary schools in italy. Technical report, IRVAPP Progress Report Series, 2010.

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

M. Bratti, D. Checchi, and A. Filippin. Territorial differences in italian students' mathematical competences: evidence from pisa. *Giornale degli Economisti e Annali di Economia*, 66(3):299–335, 2007.

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

A. Cavalli. *Insegnare oggi. Prima indagine Iard sulle condizioni di vita e di lavoro nella scuola italiana.* Bologna, Il Mulino, 1992.

A. Cavalli. *Gli insegnanti nella scuola che cambia. Seconda indagine IARD sulle condizioni di vita e di lavoro nella scuola italiana.* Bologna, Il Mulino, 2000.

A. Cavalli and G. Argentin. *Gli insegnanti italiani: come cambia il modo di fare scuola. Terza indagine dell'Istituto IARD sulle condizioni di vita degli insegnanti italiani.* Bologna, Il Mulino, 2010.

L. Darling-Hammond. Teacher quality and student achievement: A review of state policy evidence. 2000.

M.S. Garet, A.C. Porter, L. Desimone, B.F. Birman, and K.S. Yoon. What makes professional development effective? results from a national sample of teachers. *American educational research journal*, 38(4):915–945, 2001.

M.S. Garet, Institute of Education Sciences (US), National Center for Education Evaluation, and Regional Assistance (US). *The impact of two professional development interventions on early reading instruction and achievement.* National Center for Education Evaluation and Regional Assistance, 2008.

M.S. Garet, A.J. Wayne, F. Stancavage, J. Taylor, K. Walters, M. Song, S. Brown, S. Hurlburt, P. Zhu, S. Sepanik, et al. Middle school mathematics professional development impact study: Findings after the first year of implementation. ncee 2010-4009. *National Center for Education Evaluation and Regional Assistance*, page 208, 2010.

D. Goldhaber. Mystery of good teaching: the evidence shows that good teachers make a clear difference in student achievement. the problem is that we don't really know what makes a good teacher.(feature). *Education Next (Feature)*, 1:50–55, 2002.

D. Goldhaber. Everyones doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4):765–794, 2007.

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

J. Heckman and P. Carneiro. Human capital policy. Technical report, National Bureau of Economic Research, 2003.

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

C.C. Johnson, J.B. Kahle, and J.D. Fargo. A study of the effect of sustained, whole-school professional development on student achievement in science. *Journal of Research in Science Teaching*, 44(6):775–786, 2007.

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

V. Lavy and A. Schlosser. Targeted remedial education for under-performing teenagers: costs and benefits. Technical report, National Bureau of Economic Research, 2004.

R. Santagata, N. Kersting, K.B. Givvin, and J.W. Stigler. Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students mathematics learning. *Journal of Research on Educational Effectiveness*, 4(1):1–24, 2010.

A. Thurston, D. Christie, C.J. Howe, A. Tolmie, and K.J. Topping. Effects of continuing professional development on group work practices in scottish primary schools. *Journal of In-service Education*, 34(3): 263–282, 2008.