

# Preliminary Evidence from the M@t.abel Teacher Professional Development Program in Italy<sup>1</sup>

Gianluca Argentin, Università di Milano Bicocca, Italy, gianluca.argin@unimib.it

Aline Pennisi, Ministero dell'Economia e delle Finanze, aline.pennisi@tesoro.it

Daniele Vidoni, INVALSI, daniele.vidoni@invalsi.it

Giovanni Abbiati, Università degli Studi di Milano, giovanni.abbiati@unimi.it

Andrea Caputo, INVALSI, andrea.caputo@invalsi.it

## Abstract

*In this study a randomized controlled trial is used to assess the effects of a professional development program for math teachers on student achievement and teacher practice. The experiment involves 174 lower secondary schools (6<sup>th</sup>-8<sup>th</sup> grade) in Italy's lowest performing regions, a context calling for effective policies to be identified and lacking rigorous evidence. Alongside national standard math assessments, the project collected a wide amount of information on students, on teachers and on schools. Preliminary findings suggest the program had no significant impact on student math scores during the first year (when the program was held). However, some effects on teachers' practice and on student attitudes towards math do appear. These results could be promising for improvements on student learning in subsequent years.*

Keywords: teacher professional development; randomized experiment; math achievement

JEL classification: I21; C21; C26

---

<sup>1</sup> This project is carried out with EU funding - PON Istruzione 2007-2013 (Valutazione Matabel Plus I-3-FSE-2009-2). Special thanks to Alberto Martini (who contributed to set up the project), Piero Cipollone, Annamaria Leuzzi (Ministry of Education), Annamaria Fichera (Ministry of Education), Alessandro Cavalli, Maria Pia Perelli d'Argenzio, Enrico Rettore, Jaap Scheerens and for their support and precious advice.

## **1. Aim of the paper**

In many countries the drive to improve education has triggered a season of rigorous research on what kind of instructional practices, curricula and interventions work. Italy is still lagging behind for several reasons: data on student achievement are limited and recent, there is widespread aversion to testing and little tradition of evidence-based policy evaluation. Given the relative weakness of Italian students in international assessments on mathematics and science (i.e. IEA, TIMSS and OECD PISA), there has been a recent boost in the initiatives to help schools and teachers to improve student achievements (also thanks to EU funding) and an urge to understand their effectiveness. Educational research today clearly agrees on the fact that teachers do have a fundamental influence on student results (Scheerens, 2000; OECD, 2009) and are crucial to improve students' achievement (Rivkin et al., 2005). Notwithstanding the influence of factors such as socio-economic background, family and school context, student learning is influenced by what and how teachers teach.

The average age of Italian teachers is among the oldest in Europe and they do not necessarily have formal teaching skills. Moreover, the recruitment of teachers is centrally-based and allocation to schools and teacher mobility are driven more by seniority and the position they have acquired in provincial lists than on the school's choice. High teacher turnover among schools leads to segregation of the more experienced teachers within some schools, where students reach better performances (Barbieri et al., 2010). In this context schools have little lever to renew the teacher body and are rather drawn to focus on how to improve the existing staff. These features suggest that investing in teacher professional development could be a way to increase the effectiveness of the school system. Renewing the pedagogical

skills of the old and not specifically trained labor force and reinforcing teacher communities within schools could be two important leverages to increase students achievement. Improvement for teachers already in service might be sought through training following different approaches: helping them to understand more deeply not only the content they teach and the ways students learn, but also providing alternative solutions, methods and materials to present the contents.

This study investigates the effects of a specific teacher professional development program called [M@t.abel](#)<sup>3</sup>. The program, supported by the Italian Ministry of Education, covers a substantial fraction of the lower secondary school curriculum<sup>4</sup>. It is becoming popular among teachers and spreading around the country. M@t.abel is being eagerly promoted in the four regions of Southern Italy, thanks to European Union funding.

According to international and national testing, regions of Southern Italy show the lowest levels of math achievement. The OECD-PISA 2009 findings reveal that one student out of three is unable to properly master most elementary and routine tasks in mathematics (the ratio is only one out of ten in the North)<sup>5</sup>; TIMSS shows that while performing above the international average on 4th grade in math, there is a significant worsening of results among 8th graders, placing Italy among the poor performing countries. In the 2010 INVALSI<sup>6</sup> national assessment on 6<sup>th</sup> grade students, the share of correct answers in Southern Italy was on average 4

---

<sup>2</sup> M@t.abel is an acronym for *Matematica, apprendimenti di base con e-learning*. We evaluate here the most recent version of this program, carried out with EU funding - PON Istruzione 2007-2013.

<sup>3</sup> M@t.abel is an acronym for *Matematica, apprendimenti di base con e-learning*. We evaluate here the most recent version of this program, carried out with EU funding - PON Istruzione 2007-2013.

<sup>4</sup> There also is an upper secondary version of the program. However we limit this study to lower secondary school, which has homogenous curriculum.

<sup>5</sup> OECD-PISA 2009. We refer to the share of 15-year old students with at most competence level 1 in mathematics.

<sup>6</sup> INVALSI (Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione) is a research institute of the Ministry of education holding the specific aim to evaluate Italia schools. Among other tasks, a recent law prescribed that the INVALSI should carry out standardized measurement on students' ability in reading and math at various levels of education.

percentage points less than in Italy as a whole. The available empirical evidence suggests that the differences in student performance between Northern and Southern Italy cumulate over time. Whilst limited in primary school, the gap gets bigger in lower secondary school.

Using a randomized control trial, we seek to detect whether the M@t.abel program makes a measurable difference in promoting student achievement and attitudes and modifying teaching practices. To our knowledge, this is a totally new attempt in the Italian school system<sup>7</sup>. It is performed on a large scale program and not through a pilot study. The experiment meets standard requirements for the identification of rigorous evidence in the field of education<sup>8</sup>.

In this paper we present the effects estimated at the end of the first year of the experiment. The study is continuing with a longitudinal sample collecting data on students and teachers for another two years.

The paper is organized as follows. In section 2 we briefly describe relevant literature; in section 3 the Italian school context and the M@t.abel program; in section 4 we illustrate the design of the experiment and the data collected; in section 5 we show the effects of the M@t.abel program on student achievement and attitudes and, in section 6, on teacher practice.

---

<sup>7</sup> In recent years, a random assignment evaluation was carried out to test the effectiveness of chess extra-curricular classes on mathematics achievement (Argentin et al., 2012).

<sup>8</sup> Such as those in the checklist provided by the US Institute of Education Science ([ies.ed.gov/ncee/pubs/evidence\\_based/appendix\\_b.asp](http://ies.ed.gov/ncee/pubs/evidence_based/appendix_b.asp)) and the standards required by the What Works Clearing House (<http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19>).

## **2. Literature review on professional development effectiveness**

Teacher training initiatives vary widely and there is little empirical evidence on factors affecting both teaching practice and student achievement (Guskey, 2003; Fraser et al., 2007).

Previous research on in-service teacher training stresses the importance of extended duration, content-focus and peer collaboration. With US cross-section data, Garet et al. (2001) identified focus on content knowledge, opportunities for active learning and coherence with other learning activities as key features for successful training programs. These elements are confirmed by a further analyses based on longitudinal data (Desimone et al., 2002). Other researchers suggest that teaching practice and student performance are likely to improve when professional development is focused on academic content, based on teachers' collective participation and administered through long-term activities rather than one-day generic workshops (e.g., Kennedy, 1998; Ingvarson et al., 2005; Timperley, 2008).

This literature mostly contains large and small-scale studies based on teacher surveys and identifies good practices. The results of these studies are, however, seriously challenged by self selection issues, since even with longitudinal data, it is not possible to fully rule endogeneity problems out of the estimation procedure. To state it in other terms, it's hard to answer the following question: are the trained teachers performing better because they followed the training or because they are self-selected among the most able or the most motivated?

Therefore, caution must be used in the interpretation of observational evidence on this matter. More recently, some experimental evidence has been gathered on the effects of teacher training, with contradictory results (Yoon et al., 2007; Garet et al., 2010; 2011; Santagata et al., 2010).

Cohen and Hill (2000) examined teachers who participated in initiatives specifically targeted to improving the math curriculum in California. Their students scored higher on a test of the math concepts imparted by the new curriculum (compared to other initiatives).

More needs to be explored in order to understand how to shape teacher training to be effective and how to involve teachers in professional development programs. Finally, most studies concern the US providing little evidence for other countries; up to nowadays, none produced evidence for Italy.

Considering the existing evidence, [M@t.abel](#) seems a promising training program, as it holds features considered of success according to international literature. It works both on teacher practice and on math content, offering specific teaching materials and promoting group work in the classroom. It lasts the whole school year, and, finally, it involves teachers and their continuous interaction through an online virtual community.

### 3. The M@t.abel program

Promoting effective professional development programs for math teachers is challenging all around the world. In Italy programs face additional challenges, including having to address the oldest lower secondary teachers in the world<sup>9</sup>, low wage differentiation (OECD 2007) and a recruitment system which does not imply specific training in teaching. In the Italian educational system, there is no formal teachers' assessment and no differentiation in the career pathways. Teachers declare more frequently than in other countries that they miss feedback about their job (OECD 2009). Moreover, most math teachers did not graduate in math or physics and should be probably considered as out-of field teachers<sup>10</sup>.

Although in-service training is required, schools have little resources to actually carry out such programs. Incentives to attend professional development are few at school and individual level.

The M@t.abel program evaluated in this study aims at increasing lower secondary school math achievement in four Southern regions of Italy<sup>11</sup>, providing teachers with alternative solutions and methods for presenting traditional contents. The main idea is that students, rather than learning abstract formulas and ideas, should be engaged in solving real life problems through mathematical tools and concepts. The program is addressed to tenured<sup>12</sup> math teachers in grades 6-8 (middle school) and 9-10 (first two years in high school). It is based on formal and on-line tutoring and it lasts a full school year. There is a repository of teaching materials facing different math concepts by adopting a problem-solving perspective. Teachers are required to use at least four of these teaching materials (precisely one per major math content area) in

---

<sup>9</sup> In our sample, the average age is 52.

<sup>10</sup> In our sample, 83% graduated in other disciplines.

<sup>11</sup> The four regions are: Sicily, Apulia, Calabria and Campania. These regions cover about 85% of the Southern lower secondary student population.

<sup>12</sup> The high mobility rate of non-tenured teachers generates uncertainty and hinders the development of professional communities - which is part of what the training program tries to do.

their classrooms and to report on the experience to their tutor and peers through a structured diary. Moreover, the program encourages a virtual community of teachers to exchange views through on-line forums and discussion groups, also from home<sup>13</sup>.

Schools and single teachers within schools enroll on a voluntary basis to the program. While registering, they also indicate their preferred location, among the ones in their areas delivering the formal training sessions of the program. The delivering takes place through selected schools (called "*presidii*") with proper facilities for tutors and teachers meetings. Although a substantial part of the training is done at distance, through an e-learning platform, M@t.abel actually starts off if at least 12 teachers sign up for the same location.

Since most professional development programs are undertaken on a voluntary basis (and this is usually the case in Italy) an additional problem arises in analyzing them. Indeed, teacher's enthusiasm, motivation or other non-observed factors that might influence student achievement independently from the attended training (Murnane, 1985). For those teachers who actually sign up for a professional development program of extended duration, two further factors must be taken into consideration: a) actual completion of the training and b) self selection at school and individual levels which can lead to under-treating precisely the low performers. We kept these considerations in mind while designing the experiment and we will discuss their implications further.

#### **4. Experiment design, implementation and data**

This is a large scale randomized controlled trial, involving 174 schools, 666 teachers and roughly 11,000 students<sup>14</sup>. It was designed as a three-year experiment starting

---

<sup>13</sup> The [M@t.abel](#) platform and tools are made available only to the teachers following the training program.



in 2009/10 and was addressed only to lower secondary school teachers (grades 6-8). A large amount of primary and secondary data was collected both on teachers and their students. Our main target measure for student performance is the INVALSI *math competence score*<sup>15</sup>, but to investigate more thoroughly the process underlying the program impact we also seek effects on students' attitudes and teachers' attitudes and (self reported) practices<sup>16</sup>.

#### **4.1 Randomization and validity**

The identification strategy is based on a typical treatment-control comparison between students clustered by classes (and therefore by their teachers) and schools<sup>17</sup>. Given the importance of peer collaboration in the [M@t.abel](#) approach, only schools having enrolled at least two teachers were considered for this experiment. Schools randomly assigned to two groups: the teachers belonging to the first group received the specialized training in year 2009/10 (treatment group), those belonging to the second were delayed admission for one year (control group), thus admitted to the program in year 2010/11. We stratified the schools according to geographical criteria (namely by province, isolating the city of Naples and Palermo as specific strata) and by peer participation at the school level (schools with less than 5 teachers enrolled and schools with 5 or more), obtaining 31 non-null sample layers. Then 54 schools were randomly assigned to the control group, proportionally to the distribution of the schools in the sample layers. The remaining 120 schools were assigned to the treatment group and invited to participate in M@t.abel

---

<sup>14</sup> The experiment involved about 2.2 percent of the overall 6-8 grade student population in the four regions under study.

<sup>15</sup> While tests for the 6th grade are those already adopted by INVALSI for the yearly national assessment of student achievement carried out in May, additional items were developed for 7th and 8th graders and then linked, using Rasch analysis, to measure all students on a common scale.

<sup>16</sup> Teachers went through extensive computer-assisted telephone interviews (CATI) during the year following the M@t.abel program, allowing for a post treatment comparison on teachers.

<sup>17</sup> Given the high teacher mobility across schools, a high dropout rate from the three-year observation should be expected, but we rely on a large sample to avoid this drawback.

immediately<sup>18</sup>. We obtained a sample of 409 teachers invited to attend the program and 172 used as control group<sup>19</sup> (invited to attend the following school year).

One class to be observed during the evaluation was assigned to each teacher, among the many in which the teacher works. In order to guarantee the maximum variability between 6<sup>th</sup>, 7<sup>th</sup> and 8<sup>th</sup> grade within each school for both treatment and control group, we built a stratified random assignment of classes to teachers. Students being observed throughout this experiment are about equally distributed across the three grades and the teachers were asked to implement the [M@t.abel](#) teaching materials in the assigned class<sup>20</sup>.

Thanks to the large amount of information collected, we were able to test the equivalence between treatment and control group across a unusually wide range of variables at school, teacher and student levels (more than two hundred). Controlling for the randomization variables, we found small but statistically significant differences<sup>21</sup> only on a reduced set of factors<sup>22</sup>, probably due to the extremely wide set of variables used in the comparison between treatment and control group. By large, the internal validity of the experiment is verified. In any case, we run several

---

<sup>18</sup> The number of schools belonging to the control group is smaller than the treatment group for two reasons: to reduce the inconvenience of treatment delay as much as possible and to ensure enough room for the experiment to hold even in case of drop-outs from the treatment during the school year. Reference is made to the number of teachers participating to the randomization. Those actually observed (net of drop outs before the beginning of the program or lost data) are 172 for the control group and 410 for the treatment group.

<sup>19</sup> Initially the teachers invited to attend the program were 473 and the control group was 193. Due to uncontrollable factors (such as teachers deciding to retire or change of school at the beginning of the school year) the actual number of teachers observed was reduced: 84 had already dropped-out of the experiment, before the training actually started. The proportion of drop-out teachers was roughly the same in the treatment and control groups and their classes were not observed.

<sup>20</sup> The [M@t.abel](#) teaching materials are engaging for teachers not familiar with the approach. It was not realistic to expect that teachers would implement [M@t.abel](#) in all their classes.

<sup>21</sup> All the differences are below 10 percentage points; we adopted a very cautious definition of statistical significance ( $p$  value  $< .10$ ), considering the large number of students involved in the analyses.

<sup>22</sup> Namely: the presence of an external observer during the national math assessment. At the student level, treated and control differed on help at home with homework, parents' education and age of the 6th graders. The differences, it should be noted, were not big in size, even when statistically significant. Even if not statistically significant, the models control also for the presence of primary and secondary schools in the same institute, the altitude of the town where the school is located and town size, which show a moderate but non-significant association with the treatment.

models estimating the program effects by adding control variables to take into account these differences.

Looking at the external validity of the experiment, we detected that figures of our sample (based on self selected schools in four Southern regions) generally compare to those of the population of other schools, teachers and students in the whole of the eight regions of Southern Italy<sup>23</sup>, but not to the rest of the country.

### 3.2 Non compliance

The estimated effects could be diluted by the fact that some treated teachers did not actually carry out the whole M@t.abel program. In fact, only about 39% of the teachers can be considered compliant to the treatment protocol (table 1). The others must be considered not compliant either because they quit at the very beginning of the program (about 34%), either because they did not get the end-of-training certificate, typically because they did not use the four prescribed teaching units in the classroom (about 14%) or because they intended to participate but the training course actually never started in their area (about 13%). The rate of compliance is similarly distributed across the three school grades, ranging from 40.7 in 6<sup>th</sup> grade, to 44.0 in 7<sup>th</sup> grade and 39.6 in 8<sup>th</sup> grade. We did not observe crossover among the control group teachers<sup>24</sup>.

**Table 1 – Compliance in terms of teachers and students**

	Teachers	Students
--	----------	----------

<sup>23</sup> The only relevant difference is that the observed schools are located more frequently in urban areas and less in the mountains. Only Southern Italy is considered in our external validity tests; the divide with the rest of the country is striking and our results could be not applied to the other areas of Italy.

<sup>24</sup> Data collected directly from schools allowed us to check that there were no close substitutes to M@t.abel available to those in control group. Schools all benefited from a wide range of other of activities or projects involving math (such as after-school homework assistance, math software, etc.) but there was no significant difference between the two groups.

Intention to treat	409	7.692
Treated (compliers)	156	2.986
Control	172	3.372
Total	581	11.064

The reasons behind non compliance and the self selection of compliers have been analyzed using multivariate binary logistic models: full compliance is associated to younger age and participation to previous in-service training opportunities. In addition, personal motivation to enroll to the program, as reported by teachers themselves, is a significant predictor of compliance: teachers enrolled by their own request show a higher rate of compliance with respect both to teachers informed by school principals and to reluctant teachers directly asked to enroll by the school principal. This result does not necessarily mean that the current recruitment process, whereby school principals enroll teachers, is ineffective (indeed the vast majority of teachers are enrolled by suggestion of the school principals).

Another source of non-compliance is due to geographical elements, such as the mountainous nature of the area in which the school is located, which inhibits teachers in less-urbanized areas to participate to the formal parts of the training course. Indeed, the main reason reported by non-compliers to justify their dropping out is the distance from the course location, followed by the time-constraints. The program requires time to reach and attend the lessons; time in the classroom to use the materials with students; time to report about those experiences; digital skills to download materials and exchange comments with colleagues.

## 5. Short term effects on student math performance

The effects of one-year of the M@t.abel program on student and teachers are estimated in terms of intention-to-treat (ITT) by OLS models. Considering the non compliance rate in our sample, we also estimated the average-treatment-effect-on-the-treated (ATT), instrumenting full compliance with the assignment to the treatment, using an instrumental variable regression model and rescaling the effect. The ATT estimates are displayed despite the results regarding non compliance do not suggest that complier and non complier could be considered fully equivalent. Therefore it is not reasonable to expect that the effects obtained on compliers can be extended to non compliers.

In the following sections, we present the base models on students and teachers, which control only for the randomization stratification variables and the presence of an external observer during the national math assessment<sup>25</sup> and correcting the standard error of the estimates for the class clusterization of our data. For robustness, we run several models, using different sets of control variables, for which full equivalence between treatment and control did not hold. The results of our experiment do not change.

Looking at student math performance, we consider three target variables:

- the overall math score measured through the national assessment test and estimated using a Rasch model (scaled to an average of 500 and standard deviation of 100 for the 7<sup>th</sup> grade);
- the frequency of skipping at least one item (with the exception of the ones skipped because of not reaching the end of the test);

---

<sup>25</sup> In order to avoid cheating or manipulation, for the sake of this experiment, external observers were hired to guarantee the regularity of the national math assessment administration (covering about 83% of the observed classes).

- the frequency of not completing the assessment, not reaching one or more of the last items in order of presentation.

**Table 2 – Average impact on student math performance**

	Descriptive statistics		Effect estimates and standard error (OLS and IV regression)	
	Treatment	Control	ITT	ATT
Math score (mean)	493	496	1,8	4,8
Skipping items (%)	65,8	59,9	6,3**	16,7**
Not completing the test (%)	6,8	5,4	1,2	3,1

Note: The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant at the 1, 5, and 10 percent level.

The results reported in table 2 show that the treatment has no significant impact on the main outcome, the average math competence score<sup>26</sup>. Students in the classes where teachers participated to the M@t.abel professional development program actually present, *ceteris paribus*, an average slight advantage in the performance, although not statistically significant. However, the program seems to have an undesired impact on increasing the propensity of the treated to skip at least one item, despite this behavior did not affect the overall number of items skipped or

<sup>26</sup> The result does not change considering the median score through a quantile regression model.

double-marked, nor the propensity of not completing the assessment. This could be a short-term behavior due to the fact that treatment students were more concerned by doing well during the test and avoiding just guessing the answer. The following table, displaying the effects of [M@t.abel](#) on students' attitudes, reinforces this interpretation. In fact, students of treated teachers report less frequently the causes of academic failure to chance or to bad luck and a higher level of anxiety while taking the test. If we combine this element with the disposition to a higher responsibility towards their own academic failures, we could imagine that the treatment produced a perfectionist attitude and a preference for not answering instead of trying to guess the right answer. Another unexpected short term negative effect of [M@t.abel](#) on students lies in them reporting not having had enough time on a given subject - a feeling which is more frequent among the treated. This effect is probably due to the protocol time constraints and it matches the treated teachers complaints about the little time required to implement the M@t.abel approach in the classroom. A positive effect of the program is that we detect a stronger self confidence in math among the treated students.

All these effects are significant even if their magnitude is low. More important, these short term effects on students' attitudes could be promising for future improvements: feeling responsible for their own math achievement, feeling confident about their skills and being perfectionist in solving problems could be positive basis to learn math, especially as the teachers become more familiar with the M@t.abel approach.

**Table 3 – Average impact on student attitudes towards math and school**

<b>Dimensions</b>	<b>Value controls</b>	<b>ITT</b>	<b>ATT</b>
-------------------	---------------------------	------------	------------

*Attitudes towards math*

5 items factor (std score)	-0,05	+0,05	+0,12
4 items factor (std score)	-0,05	+0,05	+0,12
Item “In math I’m good” (1-4 points scale)	+2,78	+0,05**	+0,12**

*Curriculum pace*

We proceeded even if some classmates did not understand the topic (1-4 points scale)	+1,55	+0,07***	+0,17**
--	-------	----------	---------

*Failure attributions*

Attribution of failures to bad luck (0-6 points scale factor)	+0,19	-0,04***	-0,09***
---	-------	----------	----------

*Test Anxiety*

4 items factor (std score)	-0,04	+0,05*	+0,13*
I was so nervous I could not find the answers (1-4 points scale)	+1,89	+0,06***	+0,16***

---

Note: The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant at the 1, 5, and 10 percent level.

Given that specific sub-groups could have benefited most from the program, we explored heterogeneity of the effects among different groups of schools, teachers and student. There are some hints only about one feature: teachers’ age group could be a factor affecting the effectiveness of the intervention. Students of middle aged teachers (in our sample 50 to 55 years old) show a significant positive effect of [M@t.abel](#) on their average math score (ITT: 15,2), but the small size of teacher and school subsamples suggests cautiousness in deriving conclusions.

## 6. Short term effects on teacher attitudes and classroom practice



Being a professional development program, M@t.abel should trigger changes in teachers' knowledge, practices and attitudes<sup>27</sup>. It seems relevant to examine whether there are hints to changes in the way teachers are teaching, because these results could be promising about effects on the students in the following years: we should remember that our evaluation was conducted during the first year, when teachers were attending the professional development course. These effects were explored thanks to information collected through specific questionnaires, including batteries of questions on attitudes (attitudes towards math teaching, self-efficacy perception, job satisfaction) and instructional practices (classroom activities, evaluation of students, materials and instruments used to teach, interaction with colleagues)<sup>28</sup>.

We estimated the effects on instructional practices and attitudes using linear models on pseudo-continuous variables (expressed on a 1-10 points scale), and linear probability models on binary variables<sup>29</sup>.

Table 4 shows the questionnaire's items where we found effects robust to different model specifications, with and without controlling for non equivalent covariates. As concerns categorical variables, we considered robust only the effects present both on the average and on the dummy variables derived from the categorical ones<sup>30</sup>.

**Table 4 – Differences across treatment and control groups on teachers' attitudes and practices (ITT and ATT )**

---

<sup>27</sup> We were not in the position to assess the improvement in teacher content knowledge, although M@t.abel is a highly content oriented program.

<sup>28</sup> Both CATI were conducted in January. The post-treatment interview therefore occurred when the treatment group had concluded the professional development program and started a new school year and when the control group was starting to attend the program (only 57% of the control group actually started [Mat.@bel](#) the following year).

<sup>29</sup> Variables expressed on a ordered, non continuous scale were dichotomized in order to estimate a linear probability model. We performed this change because routines in statistical packages do not allow for instrumental variables models with ordered outcomes.

<sup>30</sup> Categorical variables were re-scaled in order to estimate the effect on the means.

Item	Value controls	ITT		ATT	
		$\beta$	SE	$\beta$	SE
<i>Interactions with colleagues<sup>a</sup></i>					
Preparation of teaching materials jointly with colleagues, 1-3 times a week ( <i>Ref: less frequently</i> )	47,2%	+12,3***	4,5	+31,5***	11,5
Exchange of point of views on how to teach a particular topic, 1-3 times a week ( <i>Ref: less frequently</i> )	25%	+7,2*	3,8	+18,4*	9,8
<i>Attitudes towards math teaching (1-10 points scale)</i>					
Many students show difficulties while doing abstract reasoning	7,04	-0,56***	0,17	-1,42***	0,46
<i>Self-efficacy perception (1-10 points scale)</i>					
Make the students work together	7,45	-0,21**	0,1	-0,54**	0,27

Note: The symbols \*\*\*, \*\*, \* indicate that coefficients are statistically significant at the 1, 5, and 10 percent level; <sup>a</sup> Coefficients of the battery "Interaction with colleagues" are expressed in percentage points.

The overall picture of the robust effects of the treatment presented in table 4 show that treated teachers differs considerably from their colleagues in the control group, despite the full equivalence on these items before the program started. In the school year following the treatment the complier teachers became significantly more eager to collaborate with peers at work, both in the preparation of didactic materials and in discussing the better way to present a concept to the students. The size of the effects seems relevant, both on ITT and ATT estimates. Since collaboration among teachers can be considered one of the effectiveness-enhancing factors (Scheerens, 2000), this effect could be particularly promising in Italy, also considering the high teacher turnover among schools.

Moreover, with respect to attitudes, treated teachers show less traditional attitudes towards teaching and they realized to be not fully effective in making students work in group.

On the whole, the treatment seems to have had some effects in the expected direction: teachers in the treatment group show a higher openness towards innovative practices and team work and report less traditional points of view on

teaching. On the other hand, the treatment seems to have led teachers to learn their limits, making them feel less confident about their effectiveness in promoting collaboration among students. A large number of other effects slightly above the threshold of significance reinforce this interpretation, showing that teachers do act in a more innovative and collaborative way and, at the same time, the treatment gave them more resources enabling them to re-think their role and their way to teach.

## **7. Conclusions and future work**

This paper studies the effects of the [M@t.abel](#) teacher training program, on students' math competence scores and attitudes, and on teacher behavior with a first ever randomized experiment in the Italian school system, on a sample of teachers of four Southern regions.

Although preliminary findings show no significant impact of the M@t.abel program on student math performance, some effects on students' attitudes and teachers' practice do appear and could be promising for student improvements in subsequent years. Students of treated teachers show a more positive attitude towards math and there are signals of a greater responsibility in learning and facing the national math assessment. Moreover, the effects found on teachers' attitudes and practices suggest there has been a change towards a more innovative way of leading the classroom, especially promoting more frequent exchanges with colleagues.

Overall, the program has not been effective in the short run, but the occurred changes in teachers and students attitudes could activate mechanisms leading to future positive effects on math achievement in the following school years. The study is continuing building up a longitudinal sample of the same students for another two

years. Thanks to this approach we will be able to detect possible effects at a distance, or as teachers gain familiarity with the M@t.abel program.

The analysis of the implementation process (especially the compliance to the treatment) provided some insights for the design of future similar programs, which should be targeted to those who might benefit more. The high rate of drop-outs, mainly associated to older age, and the time demanded to attendees both inside and outside the classroom, could question the fact that M@t.abel is addressed only to tenured – and therefore usually older – teachers.

An additional wave of schools has been recently recruited to replicate the experiment following a slightly different randomization procedure in order to decrease the dropping out, hence non compliance rates: these additional data, together with the panel component of the presented experiment, will allow further and more robust investigations. From the specifications tested in this paper, more in depth analysis should be carried out to understand whether there is a self-selection bias in schools which did not entirely collaborate to the data collection, leaving several non answered items on students' background and math grade at the end of the first term.

A more thorough understanding of how partially compliant teachers of the treatment group were distributed within the sample is also necessary to establish whether M@t.abel effects were diluted and whether the intervention should be modified to better retain those who enroll in the program in the future.

Although there remains a considerable amount of work to be done to acquire sufficient knowledge from this experiment, it is already an important step that it has been carried out. Developing ex ante the evaluation design in cooperation with the Ministry of Education and the M@t.abel Scientific Committee has in fact contributed to a substantial rationalization of the program design itself. Once a randomized experiment was agreed upon, the institutional actors leading the professional

development program were driven to streamline the activities and actually reinforce their key features in terms of duration, content-focus and peer collaboration. In practice this has meant, insuring that the program actually started in early autumn so as to allow its completion by the time the INVALSI standard math assessment would be held in May. It also implied fixing more precise and homogeneous requirements on the training, such as requiring four teaching units to be used in the classroom and encouraging school-level rather than teacher-level participation.

Before randomizing the treatment and control group and assigning each teacher a random class, there was substantial fear that teachers would complain for being excluded from the training program, disobey the recommendation on which class to practice M@t.abel and/or be unwilling to collaborate to the data collection. The actual picture was very different. Many teachers asked for explanations on the logistics of the evaluation through the web-site contact e-mail, but less than 4% refused to answer the pre-treatment interview and less than 8% the post-treatment interview. Denials were not necessarily related to “disappointed” control teachers. Moreover, teachers in the control schools allowed for their students to be assessed even when they were 7th or 8th grade (in 6th grade national standard tests are compulsory for all schools anyway). Within the treatment group only a very minor share of the teachers decided to practice M@t.abel in a class other than the randomly assigned one. We see these as signals of positive reception of the experiment, that are therefore an encouraging note for future randomized trials in education.

## References

- Allen J.D. (2005), Grades as Valid Measures of Academic Achievement of Classroom Learning, *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, v.78 n.5 p.218 May-June 2005.
- Argentin, A., Romano, B., Martini, A. (2012), Giocare a scacchi aiuta a imparare la matematica? Evidenze da una sperimentazione controllata, in Trincherò, R. (ed.), *Gli scacchi, un gioco per crescere. Sei anni di sperimentazione nella scuola primaria*, Milano, Franco Angeli.
- Bratti M., Checchi D., Filippin A. (2007), Territorial differences in Italian students' mathematical competences: evidence from PISA, *Giornale degli Economisti e Annali di Economia* 2007, 66(3): 299-335.
- Bloom H.S. (2007), "The Core Analytics of Randomized Experiments for Social Research" in Handbook of Social Research, Sage Publications.
- Cipollone, P. & Sestito, P. (2010). *Il capitale Umano*, Bologna, Il Mulino.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California, *Teachers College Record*, 102(2), 294-343.
- Dardanoni V., Modica S., Pennisi A. (2009), Grading Across Schools, *The B.E. Journal of Economic Analysis & Policy*: Vol. 9 : Iss. 1 (Topics), Article 16.
- Darling-Hammond L. (1999), *Teacher Quality and Student Achievement: A Review of State Policy Evidence*, Seattle: University of Washington, Center for the Study of Teaching and Policy.
- Fraser, C.; Kennedy, A., Reid, L. Mckinney, S. (2007), Teachers' continuing professional development: contested concepts, understandings and models, *Journal of In-service Education*, Volume 33, Number 2, June 2007 , pp. 153-169(17).
- Garet, M., Porter, A., Desimone, L. Birman, B., & Yoon, K. (2001). What makes professional development effective? Analysis of a national sample of teachers, *American Education Research Journal*, 38(4), 915-945.
- Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., et al. Doolittle, F. (2010), *Middle School Mathematics Professional Development Impact Study: Findings after the first year of implementation* (NCEE 2010-4009), Washington, DC: National Center for Education Evaluation and

- Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011), *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Guskey, Thomas R. (2003), What Makes Professional Development Effective?, *Phi Delta Kappan*, v84 n10 p748-50 Jun 2003.
- Ingvarson L., Meiers M., Beavis A. (2005), Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy, *Professional Development for Teachers and School Leaders*, Australian Council for Educational Research.
- INVALSI (2010), Sistema Nazionale di Valutazione a.s. 2009/2010, Rilevazione degli apprendimenti. Prime analisi. Parte I-II, <http://www.invalsi.it/snv0910/> .
- Kennedy, M. (1998). Form and substance of in-service teacher education (Research Monograph No. 13). Madison, WI: *National Institute for Science Education*, University of Wisconsin–Madison.
- Koretz, D. and Berends M. (2001): Changes in High School Grading Standards in Mathematics, 1982–1992, *RAND Education*, Prepared for the College Entrance Examination Board.
- Ministry of economy and finance and Ministry of education (2007), *Quaderno Bianco sulla Scuola*.
- Murnane, R. J., & Phillips, B. R. (1981, Fall). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance, *Economics of Education Review*, 1 (4), 453-465.
- OECD (2007), *Education at a Glance*, Paris: OECD.
- OECD (2009), *Creating Effective Teaching and Learning Environments - First Results from TALIS*, Paris: OECD.
- Santagata R., Kersting N., Givvin K.B., and Stigler J.W. (2011), 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*, vol.4, n. 1.

Scheerens, J. (2000), *Improving School Effectiveness. Fundamentals of Educational Planning*, Paris, Unesco.

Yoon, K. S., Duncan, T., Wen-Yu Lee, S., Scarloss, B., Shapley K. L. (2007), *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues and Answers*, Report, REL 2007 – No. 033. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest, <http://ies.ed.gov/ncee/edlabs>.