



Counterfactual impact evaluation of a pilot program for the inclusion of migrant adolescents.

*Valentina Lamonica, Elena Ragazzi and Lisa Sella**

CNR-Ircres

Via Real Collegio 30

10024 Moncalieri (TO)

*Corresponding author lisa.sella@ircres.cnr.it

Paper to be presented at the organised session SO-06 **Evaluation of social policies**

1 Introduction

The present case study describes the starting steps of a counterfactual evaluation design assessing the impact of a pilot training program addressed to non-EU migrant adolescents at high dropout risk in junior high school. Piazza dei Mestieri, a private training centre for adolescents, and some State junior high schools in Turin, Italy, jointly develop the intervention. The project involves as well some junior high schools and training centres in France (Chambery and Lion) and in Belgium (Liège), which are supposed to develop similar interventions on immigrant students in the next future. The idea of this transnational and cross-institutional partnership is to enhance the transfer of good practices.

All students involved in the program attend the second junior high class and experience problems in school integration, from a pedagogical point of view, and in the social perspective as well. This is

fundamental, since scarce participation is an important predictor for dropping out, most of all among young adolescents, and may be the basis for the development of bullying behaviour. Hence, the intervention is supposed to reduce drop out risk at school by a multifaceted treatment on the whole educative community, involving students, teachers, and families. A recent preliminary experience on Italian bully students showed positive results, although the evaluation was based only on qualitative evidence and on the subjective teachers' assessment.

The pilot experience presented in the paper will be evaluated by experimental counterfactual approach. Moreover, it is addressed to non-EU migrant students, who are more exposed to difficulties in their school paths, and so it concerns a more disadvantaged but also more homogeneous target.

During the experiment, treated students attend classes at the training centre for two days a week, instead of their standard school lessons, for a 290-hour professional training course. The course has strong work-oriented contents (bread-maker, chef's assistant ...) and aims at restoring students' interest in education by professionally oriented teaching, innovative inductive pedagogic methods, and individual mentorship. On their turn, school teachers are involved in periodic seminars, which reinforce their pedagogical approach to migrants' and bullying problems.

The pilot nature of the project and its mission of international sharing of best practices suggested the promoters to reserve a role for research and evaluation, charging an external partner for this. The evaluation activity, in particular, is expected to give a multifaceted contribution:

- Fine tuning of both the general approach and the pedagogical instruments (internal mission)
- Validation of perceived results to
 - enforce the sharing of pedagogical practices (trans-institutional and transnational mission)
 - strengthen the proposed model for future wider implementation (perpetuation mission)

The experimental approach adopted for the evaluation, in particular, may be conceived as an instrument for learning by doing in action-research experiences. It must nevertheless be underlined that its implementation interferes with the practices and interests of the different stakeholders involved in the experiment. This may cause difficulties that may induce subtle distortions or even hampering the trial accomplishment. Paragraph 3.1, while describing the implementation of the evaluation design, will explore and discuss the nature of these difficulties and the strategies implemented to mitigate them.

With the aim of project evaluation we constructed a grid of outcome variables, starting from the one developed and successfully applied in the Italian project¹ "Provaci ancora, Sam!". It will be monitored, for students belonging to both treated and control groups. The aim is to assess social and behavioural skills, alongside other variables concerning school performance.

¹ The project has been implemented for twenty years by Fondazione per la Scuola and it aims at preventing educational failure by specific intervention at school (<http://www.provaciancorasam.it/progetto>).

Previous experimental evaluations of similar programs (see Kautz et al., 2014 for a survey) provide ambiguous results: depending on program features such as treatment length, role and pervasiveness of training, tutors, mentors, families' and peers' influence, long-term analyses showed either positive (Durlak et al., 2011; Tierney et al., 1995; Orr et al., 1994) or negative (Rodrigues-Planas, 2012) effects on some outcome variables such as behavioural attitudes, character skills, educational attainment, future wages and other labour market outcomes. For this reason, the project will adopt a randomised experimental design, so as to get sound evidence on the effectiveness of the implemented activities.

The paper describes the “Oltre i Muri” project, that includes the random trial evaluation, specifying the characteristics of the target population and going into the particular pedagogical approach it adopts. The following part of the paper is devoted to the description of the evaluation exercise, justifying the evaluation design, and then discussing some relevant issues concerning the trial implementation. Finally, a description of the instruments used to derive the outcome indicators is included. These have been elaborated expressly for the project, relying on previous literature and projects, but also with the fundamental contribution of the VET operators, who represent an innovative element of the evaluation experience. Due to some administrative problems in the program kick-off, there was a 5-month delay and it was not possible to derive mid-term indicators.

2 The “Oltre i Muri” project

2.1 The target population

The evaluation exercise concerns a pilot program promoted by a non-profit foundation dealing with weak young students and early school leavers (in the rest of the paper we adopt the common acronym of ELET, early leaving from education and training). In particular, the “Oltre i Muri” project targets young migrant students included in State education courses of the lower secondary school.

In Italy, education is structured as follows:

- Primary school starting at the age of 6 and lasting 5 years;
- Lower secondary school, lasting 3 years;
- Upper secondary school, lasting 5 years, with different paths.

Attending school is compulsory until the age of 16, even though for the upper secondary one can absolve this duty also in training or in apprenticeship. It is worthwhile observing that high ELET rates are experienced in Italy, and that most of them refer to the passage from lower to upper secondary school and the following few months. For this reason, some innovative projects trying to prevent drop-out start in the lower secondary school. In this context, young migrants having recently arrived from non-EU countries suffer even greater difficulties, due to strong language barriers.

The Italian approach to the integration of young migrants is to include them in a standard education path, so as to maximise their social inclusion from the beginning. The main problem with this approach is represented by the low knowledge of the Italian language. Young foreigners on their arrival should attend non-curricular courses of Italian language (for instance classes of Italian ought to be offered by law in refugees' reception centres) but their effectiveness is low. At the same time, school teachers have not the right competencies to teach Italian as a foreign language, and have not time to supply personalised services.

The result is that language skill represents a barrier both for learning and for students' social inclusion, which was the main aim of their enrolment in general education. In these cases, teachers acknowledge reduced participation to school activities, non-satisfactory advances in knowledge and skills, and the tendency to create ethnic subgroups (in schools with high migrants' share). These in turn represent a problem, not only because they go in the opposite direction of the desired melting of cultures, but also because these groups are outbreaks of violence and bullying behaviour.

It is interesting observing that the initial project idea addressed foreign students' bullying behaviour (regardless their date of arrival in Italy and Italian skills), but the involvement of the stakeholders convinced the project manager to change the beneficiaries. In fact, teachers agreed that low language skilled students represent a major problem, that cannot be addressed properly by the school team alone, and that requires an innovative approach to prevent future problems of early school leaving and school violence.

2.2 Project description

The pilot project "Oltre i muri" is financed by the FAMI (Asylum, Migration and Integration Fund – EU funding) Programme in the context of an action intended to ease the national and transnational diffusion of best practices in the field of migrant hospitality and inclusion. For this reason, the project has a twofold aim. The first objective concerns the target population described in paragraph 2.1, and consists in the inclusion of young migrants at the social and school levels to prevent drop-out. Beyond that, the project aims to reinforce the local (school teachers' empowerment) and transnational community (best practices exchanges and design of common tools) through the exchange of good practices and the finding of effective strategies for the integration and guidance of young migrants.

The project partnership is characterized by a complex group of public and private organizations that operate in Italy, Belgium and France. The consortium leader is *Piazza dei Mestieri*, a centre addressed to youth in difficulty where vocational training is associated with continuous real working experiences. There particular attention is devoted to social inclusion policies and to preventing various forms of youth disorder, e.g. bullying and school dropout. The *Piazza dei Mestieri* experience on bullying behaviour from two previous projects, *Co.ds* (2012) and *Peer Co.ds* (2014), represents the basis on which an intervention methodology based on alternation between school and vocational training sessions has been established and successfully implemented. The "Oltre i Muri" project experimented for the first time the adoption of a similar approach to foreign students at high drop-out risk and involved in bullying activities. The good

results obtained in Co.ds and Peer Co.ds have prompted to apply the same methodology to include foreign pupils.

Two Italian junior high schools are partners of the project. The *Pacinotti* is both a primary and lower secondary school built in a multicultural neighbourhood in Turin: in the secondary school, out of 197 students enrolled, 103 are migrants. The other school partner, the *CPIA Parini*, is a particular one, since it supplies both primary and lower secondary courses to migrants of any nationality older than 16. This school is an important landmark for those young migrants, recently arrived in Italy, who do not possess a (recognised) school certificate and for this reason may not attend upper secondary schools. CPIA Parini mainly offers literacy courses and lessons to obtain the primary and secondary school certificates. These schools were involved in the project, both for the pedagogical experiment management, and for teachers' empowerment activities.

Other project partners are: *Fondazione per la Scuola*, which is a Foundation promoting innovative projects in the field of education, mainly addressing ELET issue in the middle high school; the *City of Turin; Immaginazione e Lavoro*, which is specialised in vocational education and training (VET) courses; and finally the *Research Institute on Sustainable Economic Growth (CNR - IRCrES)* in charge of the pilot evaluation.

The project also involves foreign partners, in order to address the activity concerning the international transfer of practices. They are located in Lyon and Chambéry (France), and in Liege (Belgium).

The *CEPAJ Sleat* (vocational training centre in Lyon) hosts 115 teenagers with personal, social and family problems, enrolled through either the social services or the juvenile court. The school offers different training programs tailored to students' wishes, to opportunities and to the specificity of the labour market.

The *Second chance* school based in Chambéry is characterized by the involvement of young people (aged 16 to 26) who are NEET (Not in Employment, Education or Training) but would like to start a vocational training program. This school works on reorientation of these young people, but first of all it works on skill recovery (basic knowledge as Mathematics, French, and computer) through a three step process. The first stage is the pre-engagement period, aimed at re-entering the trainee in the school and relieving its self-confidence. The second stage is characterized by an alternation among basic subjects, vocational training, and internship. The third step consists of a sixteen week internship, which is preparatory for job placement.

The last foreign partner is the *De Bouvoir* school in Liege. It represents an interesting case of inclusion of illiterate young migrants through a "catwalk class" attended for a period (from six to eighteen months), in order to ensure that the newly arrived migrants are included in traditional classes only when they have proper language skills.

2.3 The pedagogical approach

"Oltre i Muri" main objective is the social inclusion of young migrants in the school environment, in order to prevent the ELET. As previously described, the project target is composed by non EU-students with behavioural disorders (for instance with tendency to be excluded by their class

mates, with low language skills, etc.) indicated by their teachers. The specificity of the project consists in the alternation of school and training courses during the week. Through this alternation, students are stimulated to break the stalemate and they are motivated thanks to the dynamic learning context.

The intervention model elaborated by Piazza dei Mestieri is characterized by four steps. First of all, the VET centre takes in charge the student and his family. The second step concerns the development of a personal project (school guidance and special needs) elaborated by the training school and agreed with both the provenance State school and the family. The third step involves the pedagogical approach and the activities put in place to promote student' inclusion. The last step is one of the desired objectives: the development of a common project for foreign minors thanks to the exchange of good practices between Italian and foreign partners.

The most important step, which is the evaluation object, concerns the development and implementation of the pedagogical approach. Students are involved in a 290-hour path in Piazza dei Mestieri. In particular, students attend:

- 30 hours of Italian language course
- 30 hours of Mathematics course
- 30 hours of English language course
- 200 hours of laboratories (70 hours of cooking classes, 60 hours of graphic design courses and 70 hours of bartender courses).

This educative model is based on the work-related learning concept, which means that students are stimulated to learn theoretical subjects through laboratory practice. An example is the way of learning Mathematics during the cooking practice (fractions and proportions are easily understood in practical contexts). Beyond that, the pedagogical approach is individualized: each student can in fact rely on a tutor during the classes attended at Piazza dei Mestieri, whose function consists of educational support and linguistic mediation between the less literate students and the whole class. In order to promote social inclusion, Piazza dei Mestieri added peer tutors during some courses (i.e., older students from different nationalities, attending the training school, that share their experiences and encourage the inclusion process through peer learning). Thanks to these professionals and the peer tutors, students' motivation can improve, producing positive effects on school paths. This is a big issue, since the project lasts about one year and, after this period, each student ought to be able to come back to traditional school being less exposed to ELET.

During the project, Piazza dei Mestieri organizes two “experiential weeks” that involve students and their teachers in joint laboratory activity. The aim is to share with their teachers the process of change, of learning and flipping the traditional educational path (from teacher to pupils) into a non-conventional one, where the teachers can learn from pupils (e.g. how to make a coffee or a pizza in the laboratories) and students can teach to their professors. Thanks to these activities, teachers can realize their students' efforts and changes while students can touch their new skills, the same ones that their teachers do not have.

3 The evaluation design

The purpose of impact evaluation is to understand the effectiveness of the assessed actions. It will answer to the following questions:

- *Is the policy effective for the targeted beneficiaries?* (general effectiveness)
- *Are there sub-groups of beneficiaries for whom the policy is more (less) effective?* (conditioned effectiveness)
- *How large is this effect?* (impact quantification)

These questions imply the need to detect (and statistically test) a causal connection between the policy and the expected result represented by one or more objective variables.

In the hard sciences this happens by the experimental approach, which allows observing what happens if an event/action occurs and what if not. In the human sciences this is not feasible, because it is impossible to observe an individual both as receiving a given treatment and as not receiving it. So the counterfactual approach means confronting two groups of "homogeneous" people, which differ only for participation in the treatment (the laboratories, in our case study).

This homogeneity between main and counterfactual groups is very difficult to achieve. The literature identifies two possible designs:

- Quasi-experimental approach: Pairing of individuals (with proper algorithms) according to observable variables (characteristics) that are correlated with the objective variable;
- Experimental approach: creating a list of eligible candidates and then casually extracting who will receive the treatment and who not.

No alternative is exempt from difficulties or limitations. In the case of the quasi-experimental approach, the problem is that of creating this "identical" control group ex-post. In particular, in policies that affect the recovery of human deficits, unobservable variables are fundamental for individual success, and so a selection process based on observable variables (ex. gender, age, place of residence) itself can introduce a distortion.

On the other hand, in the case of the experimental approach the difficulties mainly concern the setting up of the random trial. "Oltre i Muri" project implementers express resistance, because the analytical set-up often conflicts with operative priorities. Moreover, a possible source of bias is represented by the substitution of enrolled students dropping out from the project.

We chose to rely on the experimental approach because the particular target described in par. 2.1 is largely subject to selection bias, for the important role played by unobservables. Moreover, a serious matching was hard to perform, due to the absence of profiling variables for the students. Finally, the trial sample is very small, so econometric methods to detect and correct the selection bias could not be applied.

Unfortunately, some risk of distortions remains, because some substitutions were necessary after the beginning of the program for some drop-out cases (see par. 3.1). Therefore, we decided to measure the impact on differences (dif-in-dif). This implied the necessity to measure the outcome

variables (indicators representing individual skills, see 3.2) both at the beginning and at the end of the project.

The detailed evaluation plan was agreed in a joint effort with the VET operators. In fact, the evaluation design has to rely on the object to be evaluated. In the case of random trials, this is particularly true, because the evaluation activities have to be planned from the beginning, interfering with the program organisation and the supply of services to the participants. For this reason, a successful random trial design has to take into account the instances of the policy implementers.

In our case, the definition of the evaluation plan required a long discussion concerning the criteria for the eligibility to the treatment. A major issue raised by the VET operators is that students with difficulties have to be addressed with programs conceived for their special needs, which do not often coincide with the problems manifested at school. So their point was that the programmed activities did not suit all the students recommended by the teachers. The initial plan of the research team was to randomly assign to the main and counterfactual groups all students recommended by the schools, while the unexpressed idea of the operators was to assign the non-suitable students to the control group. The process to define the evaluation plan allowed agree on a common suitable solution, i.e. defining at first a list of eligible students, then assigning them randomly to treatment or control group. This process further shortened the sample size, but it guaranteed a correct randomization.

3.1 The randomized controlled trial design: some applied issues

The intrinsic nature of the project is particularly suitable to field experiment design, which allows researchers to estimate the impact of interventions by randomized experimentation in real-world settings (Gerber and Green, 2012). In fact, the “Oltre i Muri” project investigates the effects of a pilot work-related learning intervention, which could be extended to a large number of junior high schools, if effective. Since Italian compulsory school is still far from being truly effective in foreign students’ integration, it needs complimentary support, e.g. from targeted public programs or no profit organisations initiatives. For this purpose, the pilot has been designed with a strong learning by doing mission.

In principle, since the pilot offers just a very limited number of positions (15), while eligible students are much more numerous, potential beneficiaries can be easily divided into treated and control groups by simple random assignment techniques. However, when practical issues come to the scene, the realization of the experimental design becomes much more complicated.

First of all, due to consistent administrative delay in the project approval, it was not possible to meet the very original timing: VET intervention effectively started in March, forcing operators to split the training course between two consequent school years (2016/17 and 2017/18) and excluding last-year students. Moreover, in order to speed up the pilot implementation, it became fundamental to rely on Piazza dei Mestieri network of schools, whose solid partnership has been reinforced during previous projects. Consequently, students’ selection was substantially managed by the VET operators.

These facts altered the original experimental design. On the one side, the number of students potentially involved dramatically decreased. On the other side, researchers had to work quite hard to persuade VET operators, who contacted schools and students, that participants' randomization was a feasible *modus operandi*. In fact, being Piazza dei Mestieri charged with the pilot practical implementation, it was mainly concerned with organizational issues. Treated students were supposed to attend classes at Piazza two days a week during school days, from March to November 2017. Hence, VET tutors were much worried about class arrangement, aiming at building a balanced group and avoiding the most difficult students. At the beginning, preliminary fact-finding meetings with students and their parents, actually aimed at assessing their interest in the project, were conceived by VET operators as an opportunity to cream-skimming training participants, potentially causing selection bias.

In order to avoid this bias, the agreed solution had to accommodate various aspects:

- Piazza dei Mestieri worked within its consolidate school network and all partner schools provided at least one treated student, in order to strengthen their commitment;
- Eligible students were recommended by teachers, since no students' micro-data are available;
- Eligible students from CPIA Parini were included in the sample to reach a fair numerosity both in the main and control group, even though their homogeneity with the other students is low;
- Preliminary meetings cannot be used as a cream-skimming tool, but help at identifying students that are eligible for the project;
- The main and counterfactual groups should be randomly assigned from the list of eligible students created thanks to the preliminary meetings;
- Non-random substitution has to be avoided.

Following the above criteria, the main and counterfactual groups were selected by block random assignment, controlling for both gender and age. In fact, on the one side education and behavioural programs generally show different gender effects (Kautz *et al.*, 2014), while age has to be controlled since many recommended students are older than regular second-year junior high school students (12 years old), possibly affecting final outcomes.

As explained, it was not possible to meet the original project requirement, consisting of 15 treated and 30 controls. On the whole, 43 second-year students were recommended by 10 junior high schools in the network. Preliminary fact-finding meetings excluded 8 students, who either were not interested in the project or were judged too problematic by the VET operators and so a group activity was not suitable for them. Then, eligible students were randomly assigned to the treated and control groups by tossing a coin: females were equally distributed and average group age was controlled for. However, during the first month activity, two students moved their house, dropping out of the project. In one case, the VET operators forced a non-random substitution: consequently, the record was excluded from the analysis.

Table 1– Recommended students' distribution by gender.

	Females	Males	Tot
--	---------	-------	-----

Record type	AV	%	AV	%	AV	%	Average age
Non treated	2	11,1	16	88,9	18	100,0	15,7
Treated	3	21,4	11	78,6	14	100,0	15,1
Excluded (not eligible)	3		5		8		
Drop-out			2		2		
Non-random substitution	1				1		
Tot	9		34		43		

Table 1 illustrates the final distribution of the treated and control groups by gender. Students in the treated group are 14, since the non-random substitution was excluded. Controls are just 18. Numbers are very limited, but randomization guarantees full inter-group comparability.

3.2 Output variables and the measurement of performance indicators

The project evaluation is aimed at understanding the general effectiveness of the project (in terms of its ability to reach the target) and its real impact. As mentioned before, the main objective of “Oltre i muri” is the *social inclusion* of non-EU young students. To test the achievement of this goal, the impact evaluation investigated two types of indicators, one deriving from students’ auto-evaluation, the other from teachers’ judgment (hetero-evaluation). The second goal concerns *learning of basic subjects*. The evaluator aims at understanding the difference between students’ initial and final level of knowledge in Italian, English, and Mathematics.

Finding valid indicators to catch both dimensions - social inclusion and educational learning - was quite hard because of the experimental nature of the project. Indicators and test instruments, rooted in the sociological and psychological literature, have been conceived in close connection with the teachers and educators of Piazza dei Mestieri. The tests have been administered to both the factual and counterfactual groups during the same period, in order to ensure comparability. At present the project is in progress: we collected the entrance tests, but final tests will be administered in October 2017, at the end of the project.

Tests for *social inclusion* were developed by means of two different instruments: a hetero-evaluation filled by teachers and a self-evaluation filled by students. For the hetero-evaluation we used an instrument validated in an Italian project, aimed at preventing school dropout and focused on the relationship with peers, with adults and, more in general, on behavior at school. The teachers filled the test, providing an initial idea about social inclusion and behavior that could lead to students’ social exclusion. The self-evaluation was developed using a pictographic questionnaire tested for the evaluation of soft skills (Maćkiewicz and Ciecuch, 2016; Barbaranelli et al., 2003; Heckman and Kautz, 2012). Notwithstanding the questionnaire was conceived for children aged 7- 10, it seemed a good solution to overcome students’ language barriers. The main feature of the pictographic questionnaire is the mix between written sentences and images that can facilitate the comprehension. The questionnaire aims to identify five behavioral dimensions

that are considered predictors of social inclusion: extraversion, agreeableness, conscientiousness, openness, neuroticism (Maćkiewicz and Ciecuch, 2016; Ahadi and Rothbart, 1994).

For testing *learning*, three questionnaires in the base subjects (Italian, English, Mathematics) have been designed by the VT teachers involved in the project. Both the Italian and English tests were designed in order to assess the level of language comprehension in a communicative context. The tests were calibrated on the A1 / pre-A2 level of the Common European Framework of Reference for Languages. The questionnaires were composed of seven exercises, each one containing a different number of items and weights, in order to create a skills scale (i.e., higher values assigned to more difficult items). The English test included little use of Italian terms (the instructions on the exercises were available both in Italian and English), since the logic is to assess the linguistic competence in English, which could be high in a migrant students, even when its competence in Italian is low. While the language tests have been created with the same theoretical logic, the mathematics test was conceived in a rather different way, in order to investigate the basic math skills (the four operations: addition, subtraction, multiplication, division, and basic geometry) described in both Italian and English. Most exercises were purely based on ideograms and so didn't need any linguistic skill.

Data collected by these tests will be analyzed in combination with objective data collected at school, e.g. school failures, good behavior and grades, in order to have a complete picture of school inclusion issues.

4 Preliminary results from entrance tests

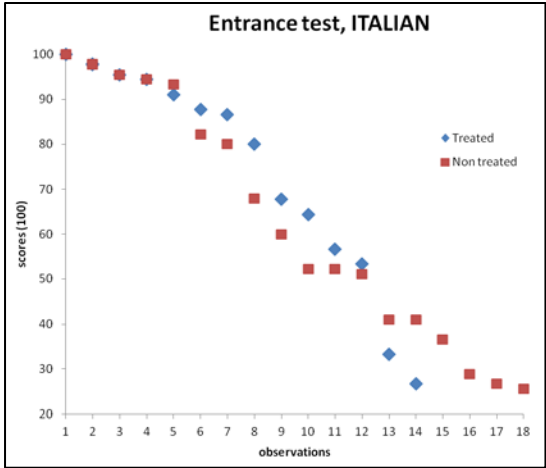
This section contains some preliminary results about entrance tests on basic subjects. Since selection into treatment is random, differences in test performance between the main and control groups are not due to selection bias, rather, they can be attributed to small sample bias or to procedural problems.

Due to project delays (see sec. 3), entrance tests could not be administered neither at the very beginning of the project, nor in a fully homogeneous setting. In fact, treated students filled the tests during their VT classes at Piazza dei Mestieri, being probably stimulated by the presence of their new VT teachers and by their involvement in a new exciting experience. On the contrary, non-treated students filled the tests during an extra-curricular visit and afternoon break at Piazza dei Mestieri. Hence, it is highly probable that these settings provide different stimuli, influencing students' test performance.

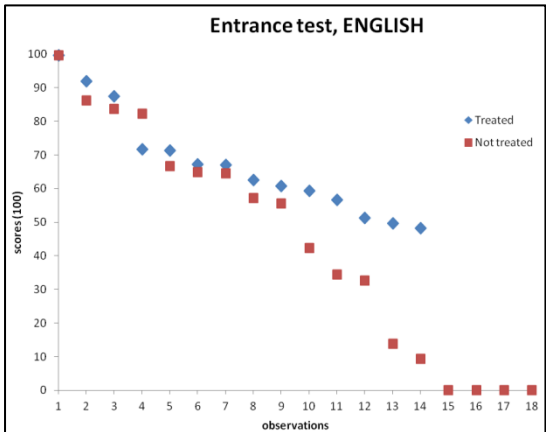
Analyzing score results, treated students perform averagely better in all basic entrance tests, showing also lower score dispersion, i.e. hypothetical more homogeneous knowledge levels. In fact, blue scores for treated students in Figure 1 are in general higher than red scores for counterfactuals. The lowest dispersion is observed in Math tests, where treated students show a very high performance on average (92 out of 100) and just two scores are lower than 90, while the counterfactuals distinguish in a middle- and an high-performance groups. However, this suggests that the test was unable to properly distinguish different knowledge levels because of its excessive simplicity. On the contrary, some counterfactual students performed very low in English test (even

four tests had a 0 score), that shows a poor average score (44 out of 100): possibly an excessive use of the Italian language in test information represented an important barrier, notwithstanding help examples were provided. However, it is important to observe that average results in the main and counterfactual samples are much more similar for students performing at least the passing grade (60).

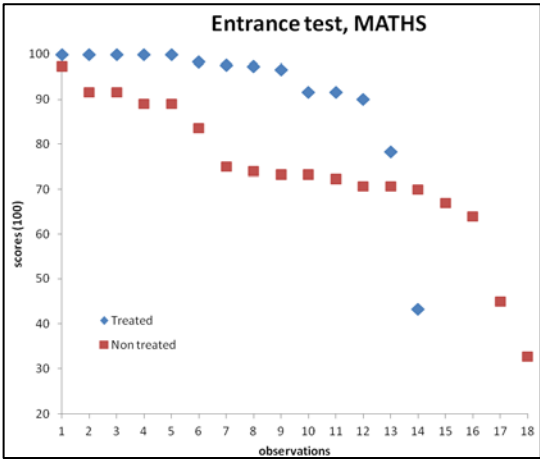
Figure 1: Entrance test scores in Italian, English, and Mathematics and corresponding statistics. *90% confidence level; **95% c.l.; ***99% c.l.



	ITALIAN TEST	
	Main	Control
Obs.	14	18
Average	74,0	62,6
Std. dev.	24,1	26,7
Max	100	100
Min	26,7	25,6
obs >=60	10	9
obs <40	2	4
obs >=80	8	7
Avg.>=60	86,6	85,7
T-test	0,608	



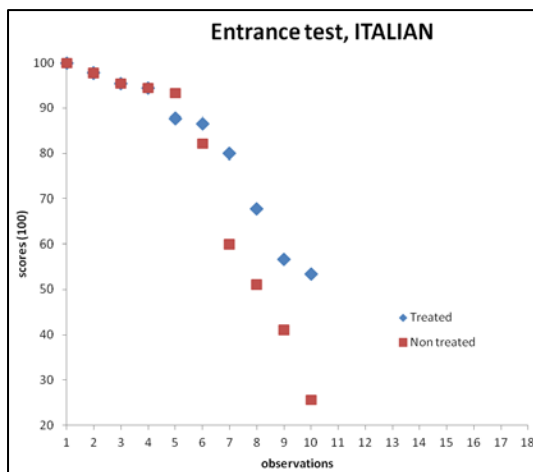
	ENGLISH TEST	
	Main	Control
Obs.	14	18
Average	67,4	44,0
Std. dev.	15,8	34,0
Max	99,6	99,6
Min	48,2	0
obs >=60	9	7
obs <40	0	8
obs >=80	3	4
Avg.>=60	75,4	78,2
T-test	0,074*	



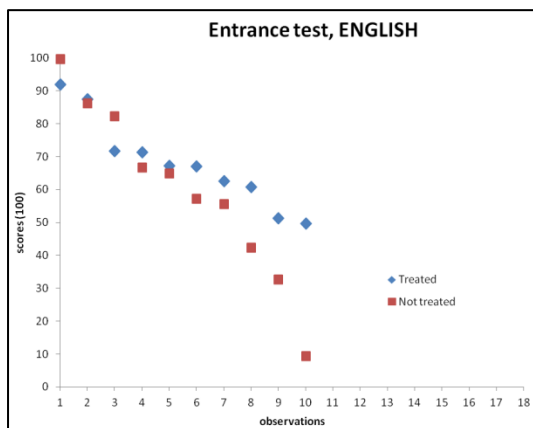
	MATHS TEST	
	Main	Control
Obs.	14	18
Average	91,8	73,9
Std. dev.	15,2	16,2
Max	100	97,3
Min	43,3	32,7
obs >=60	13	16
obs <40	0	1
obs >=80	12	6
Avg.>=60	95,5	78,3
T-test	0,019**	

Finally, the Italian test seems the most balanced one, detecting different knowledge levels. However, lower competences emerge among the counterfactuals (50% students had no passing grade, while among the treated they were just 28%), partially explaining the English test results too. In accordance, mean comparison T-tests show that main and counterfactual distributions are statistically different in the case of both English and Math tests, while no significant difference is detected in Italian test.

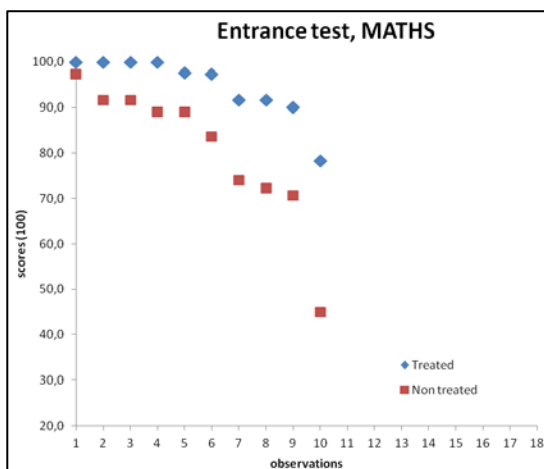
Figure 2: Entrance test scores for non-CPIA students and corresponding statistics. . *90% confidence level; **95% c.l.; ***99% c.l.



	ITALIAN TEST	
	Main	Control
Obs.	10	10
Average	82,0	74,1
Std. dev.	17,1	27,3
Max	100	100
Min	53,4	25,6
obs >=60	8	7
obs <40	0	1
obs >=80	7	6
Avg.>=60	88,8	89,0
T-test	0,424	



	ENGLISH TEST	
	Main	Control
Obs.	10	10
Average	68,0	59,6
Std. dev.	13,6	26,9
Max	91,8	99,6
Min	49,6	9,2
obs >=60	8	5
obs <40	0	2
obs >=80	2	3
Avg.>=60	72,4	79,9
T-test	0,455	



	MATHS TEST	
	Main	Control
Obs.	10	10
Average	94,7	80,4
Std. dev.	7,0	15,5
Max	100	97,3
Min	78,3	45,0
obs >=60	10	9
obs <40	0	0
obs >=80	9	6
Avg.>=60	94,7	84,4
T-test	0,102	

When results from CPIA students are excluded (Figure 2), main and counterfactual groups look much more similar in entrance scores. In accordance, T-tests detect no significant difference.

Finally, Figure 2 confirms that CPIA students are too heterogeneous with respect to standard school pupils, mostly because of the different institutional missions, as explained in sec. 2.2. Moreover, the operators had strong difficulties in administering entrance tests to CPIA counterfactuals, since they did not join the extra-curricular visit at Piazza dei Mestieri and tests must be directly collected at CPIA. However, CPIA classes are not truly compulsory, thus students' attendance is generally much scarce. This could be an important source of commitment bias: results are much cleaner if CPIA students are excluded from both samples. For this reason, our analysis concentrates on non-CPIA pupils.

Table 2 – Pearson's pairwise correlation between entrance test scores for non-CPIA students.

Test	Italian	English	Mathematics
Italian	1,00		
English	0,45**	1,00	
Mathematics	0,55**	0,22	1,00

*90% confidence level; **95% c.l.; ***99% c.l.

Table 2 shows Pearson's pairwise correlation coefficients between entrance test scores. It emerges that Italian results are significantly but moderately correlated with other tests, while no relation emerges between English and Mathematics scores.

5 Summary of preliminary conclusions

This paper discusses the evaluation experience of a project aimed at a better school integration of young migrants, which is still in progress. For this reason, it is currently not possible to show the results of impact evaluation, and final considerations are limited to the lessons deriving from the evaluation design implementation.

Project effectiveness will be evaluated by means of an experimental approach. We didn't rely on quasi experimental techniques for many reasons, including the small size of the trial, the lack of suitable variables for matching, the important role played by unobservables in policies that address disadvantaged people.

On the other hand, it is well known that random trial implementation may be difficult, because it has to be planned from the beginning of the project, it may face declared opposition or unexpressed resistance by the operators implementing the policy, and it may still suffer of selection bias in the case of non-random substitution of treated individuals.

We may acknowledge that all these situations occurred and they had to be properly faced. There only case of non-random substitution was excluded from the evaluation sample.

But the most interesting lessons derive from the other issues, i.e. planning/timing and resistance. Starting discussing the initial implementation, the counterfactual evaluation has been included in the project because its promoters showed an high commitment in understanding the effectiveness

of a very experimental approach to ELET. In its initial design, the project should have lasted one school year, including a fair time to implement the random trial before the beginning of the teaching activities. Due to problems in the program management by the Ministry, there was a 5-month delay in the kick-off.

	Sept	Oct	Nov	Dec	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec	...
School year																	
Scheduled timesheet	K.O.	Trial design				Laboratories				Vacation time			Evaluation				
Actual timesheet						K.O.	Trial design						Laboratories				
								Laboratories						Evaluation			

This fact caused a number of practical implications. For example, it was impossible to include in the trial last-year students, who would not have been able to attend laboratories in the next school year, so sharply reducing the number of candidates. Moreover, the project was designed so as to follow the school calendar, with a lot of advantages for both the project itself, and the evaluation task. For example, the possibility to rely on school notes as an outcome indicator is sharply reduced, because in the actual timesheet it is hard observe an impact at the end of the first cycle of laboratories (June 2017), while at end of the project final school reports are not available. Activities for a medium term evaluation of the project effects on school performance in June 2018 are possible, but the cost will not be covered.

Finally, the most impacting consequences were suffered by the random trial design. In fact, preliminary implementation activities suffered a hard rush-up, since teaching activities (laboratories) may not be compressed. As a consequence, it was difficult to select in a short time a fair number of candidates, thus reducing the homogeneity of the selected target (e.g., by the inclusion of CPIA students). Moreover, both selection and enrolment procedures had to heavily rely on the promoter, so the evaluator could not supervise all steps. Finally, some activities that should have occurred sequentially, actually overlapped (e.g., entrance tests administered during laboratory classes). Unfortunately, we observed some effects on the results of entrance tests (see section 4). Since they were administered after selection procedures, different students' commitment induced a small positive bias in favour of the treated individuals. This inconvenience will be mitigated by the diff-in-diff method, which estimates changes in the objective variables, measured both at the beginning and at the end of the treatment.

To conclude, timing is fundamental. This is often true in policy evaluation (e.g., policies addressed to firms that are subject to business cycle), but it is crucial in educational policies, which have to coordinate with the school calendar. If this fails, both the project effectiveness and the evaluation feasibility are threatened.

As far as the relationship with implementing actors, we received full support by the top managers, who agreed on the initial design, i.e. the random assignment to main and counterfactual groups of all students recommended by the schools. On the contrary, VET operators had the unexpressed idea to assign non-suitable students to the control group, causing a serious selection bias. In general, they refused to include whatever candidate without a preliminary selection of the most suitable ones and insisted on creating a well-balanced treatment group. The final solution, that

was suitable for both the operators and the evaluators, was to define at first a list of eligible students, then to assign them randomly to either treatment or control groups. This procedure reduced the sample size, but ensured both the evaluation reliability, and project feasibility.

All above considerations prove that a serious preliminary work aiming at creating a common agreement on objectives and procedures is fundamental in random trial evaluation. This common wisdom is particularly true when the client of the evaluation service is a non-government organisation, as in this case, which cannot be forced to adapt its empirical strategy. However, the preliminary brain-storming step is in the evaluator interest as well: in fact, social innovators generally need non-conventional evaluation strategies, because they generally produce very innovative and advanced experiences.

Since the target population of pilot projects is often fairly unknown (few available statistics), a great preliminary work is also needed to specially design the measurement tools. Our experience, that merged researchers', teachers' and psychological counsellors' competencies, shows that it remains very difficult to correctly calibrate the instruments for the measurement of the output/outcome variables. Both time and budget have to be invested in these activities.

In the specific case of young migrants with scarce school performance and social inclusion, we learnt that they are a very heterogeneous group. Some are fairly skilled, but it is difficult to detect their skills for both the linguistic barrier and the low social inclusion. The entrance tests had to include a much wider range of levels than firstly expected regarding teachers' files. Their judgement in this case failed, because of the scarce integration of their pupils.

Another important lesson concerns the influence of the environment where the student takes the test. We agreed a common procedure, and common instructions to be read before the test, but the entrance results are different for the main and control groups, also because the test were administered in different environments. Commitment problems have played a major role too. For this reason, it is important that tests occur before students' selection in the two groups. For practical reasons, this proved impossible and caused enormous problems of motivation in the students of the control group.

References

- Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences*, 34(4), 645-664.
- Devaux-Spatarakis, A. (2014). L'expérimentation « tellequ'elle se fait » : leçons de trois expérimentations par assignation aléatoire, *Formation emploi*, 126, Avril-Juin, 17-38. <http://formationemploi.revues.org/4180>.
- Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate behavioral research*, 44(3), 362-388.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*, 82(1), 405-432.
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour economics*, 19(4), 451-464.
- Kautz, T., Heckman, J. J., Diris, R., TerWeel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (No. w20749). National Bureau of Economic Research.
- Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth*, 55(1), 128-165.
- Maćkiewicz, M., & Ciecuch, J. (2016). Pictorial Personality Traits Questionnaire for Children (PPTQ-C)—A New Measure of Children's Personality Traits. *Frontiers in psychology*, 7.
- Orr, L. L., Bloom, H., Bell, S., Lin, W., Cave, G., & Doolittle, F. (1994). *The national JTPA study: Impacts, benefits, and costs of Title II-A*. Bethesda: Abt Associates.
- Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States. *American Economic Journal: Applied Economics*, 4(4), 121-139.
- Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly* (1982-), 21-39.