IRPET Istituto Regionale Programmazione Economica della Toscana

International Workshop of Applied Economics of Education

The effects of a dropout prevention program on secondary students' outcomes

Enrico Conti, Silvia Duranti, Alessandra Mattei, Fabrizia Mealli, Nicola Sciclone

enrico.conti@irpet.it silvia.duranti@irpet.it mattei@disia.unifi.it mealli@disia.unifi.it nicola.sciclone@irpet.it

Catanzaro, 22-24 giugno 2014

The paper in a nutshell

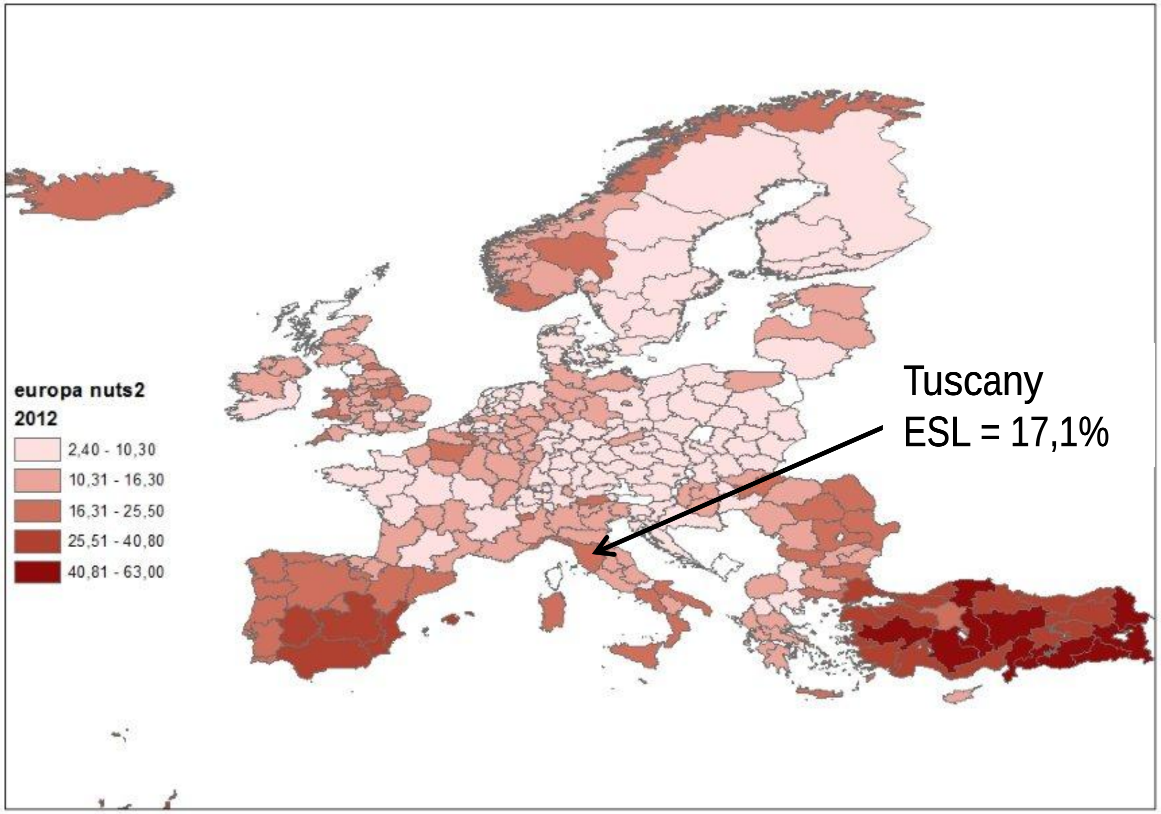
- **Objective.** To examine and validate the effectiveness of INNOVARE, a teacher-based dropout prevention program
- **Evaluation design.**
  - A quantitative statistical approach
  - A qualitative evaluation through a Focus Group
- **Data sources.** Administrative sources from schools involved, and a questionnaire on personal and family characteristics of students.
- **Main feature.** A cluster-level analysis and an individual-level analysis were conducted.
- **Basic outputs.** Both methods show a slight decrease in the probability to fail, to drop-out, and in the absence rate, and conversely an increase in the probability of postponement of the evaluation, linked to participation in INNOVARE

IRPET

Motivation

Early school leaving is a complex phenomenon closely linked to negative employment outcomes, social exclusion and poverty. In particular, unemployment and health are the key ingredients of the cost of school dropouts (Brunello & De Paola, 2011).

Early school leaving rates in Europe (nuts 2). 2012



Italy and Tuscany show particularly high and persistent early school leaving rates

* Early school leavers: the indicator is defined as the percentage of the population aged 18-24 with at most lower secondary education and not in further education or training.

The INNOVARE Program

Where? The project involves 18 first classes in 12 public vocational schools (IEFP) located in the Tuscan provinces of Florence, Pistoia, Lucca, Pisa and Massa Carrara. The Vocational school system is particularly affected by *drop out* problem, it is difficult to involve students in teaching programs and therefore to provide them with adequate basic skills.

What? An innovative teaching method inspired to the social research method called “Action Research”, successful in similar contexts (Kemmis & McTaggart, 1982).
It starts from the training of teachers to motivate or re-motivate both teachers themselves and students. Teachers, properly guided by tutors who are disciplinary experts in education and epistemology, become the designers of the new teaching method, characterized by an extensive use of educational workshops and by learning by doing, in a process of continuous comparison - reflection - correction of the educational practices implemented.

How? The project activity consists of 10 meetings between the expert-tutors and the teachers involved in the project, which then lead to the application of the new proposed teaching to their students during the second term of school year. The subjects considered by the experiment are: Italian, Mathematics, Foreign language, Integrated science, Physics and technology.

Evaluation design: the quantitative approach

- The INNOVARE study is a *cluster-randomized* trial where the unit of assignment is the class: 18 classes were assigned to the treatment of the new method (intervention group) and 35 classes were assigned to the control treatment.

- In a cluster-randomized trial, clusters are assigned to treatment or control, but often individuals are of interest. Thus the unit of assignment may be different from the unit of analysis.

- We conduct both:
  1. cluster-level analyses
  2. individual-level analyses

  Using the potential outcome approach to causal inference in both cases we tested the following outcomes:
  1. Percentage of failures
  2. Percentage of postponements of the evaluation
  3. Percentage of drop-outs
  4. Absence rate (%)
  5. Percentage of failures + Drop-out

IRPET

Observed Variables

Individual Background Variables	Class Variables	Outcome Variables (at the individual or class level)
Sex	Class size at the beginning of the school year	Failure
Year of birth	Class size at the beginning of the second semester	Postponement of the evaluation
Nationality	New entrants in the second semester	Drop-out
Late/not late	% drop-outs in the first semester	Absence rate (%)
% absence rate in the first semester	% absence rate in the first semester	Failure or drop-out
Average mark in the first semester	Average conduct mark in the first semester	
Parents' education level (primary education or higher)	Average mark in the first semester	
Parents' occupational status (employed, unemployed)	% foreigners	
	% males	
	% late students	
	% repeating students	
	% students with parents with a low education level	
	% students with unemployed parents	
	Teachers with a open-ended contract	
	Teachers between 30 and 50 years old	

IRPET

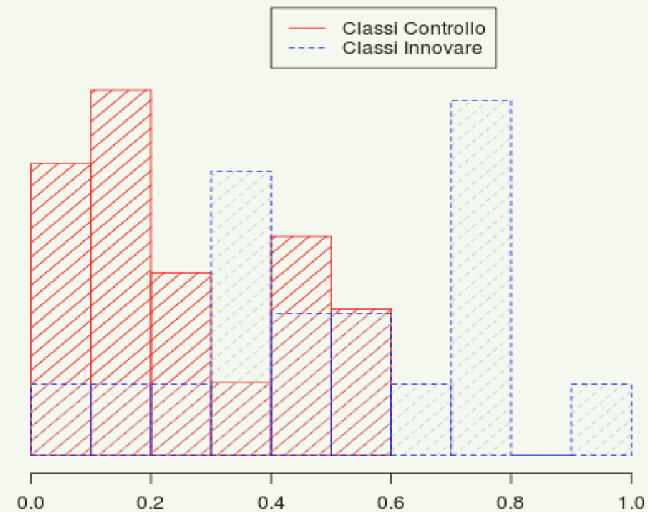
Cluster level analysis

- In Cluster-level analysis we use <u>classes as units of analysis</u>: therefore only cluster-level variables enter the analysis.
- The issue of interference between students in the same cluster does not arise, because focus is on cluster-level and we assume that students in different classes do not interfere with each other.
- The randomization inference is non-parametric in that it does not make any functional form assumption and it is exact because it does not rely on large sample approximations. Results coming out of this analysis are exact and valid irrespective of the number of group assigned to each treatment status.

A stratified cluster randomized experiment

- In order to account for differences in background pretreatment variables between the treatment group and the control group we use subclassifications on propensity score. (Rosenbaum and Rubin, 1983)
- The condition of strong “ignorability” implies that within cells defined by the pre-treatment variables the treatment is randomly assigned. So the allocation mechanism defines a *cluster-randomized* trial
- Based on the estimated propensity score, we restrict the analysis to the subsample of 48 classes that satisfies an overlap or common support condition.

Fig. Distribution of propensity score estimated between “Control” and “INNOVARE” classes



- Then, we re-estimated the propensity score using the selected subsample of 48 classes and use it for adjusting treatment comparisons for differences in background covariates using subclassification: we divided the sample into H=4 strata based on propensity score categories
- Under the assumption that data come from a stratified cluster randomized experiment, our cluster-level analysis use randomization inference to draw exact inferences for our finite population (sample) of size K=48.

IRPET

Cluster randomization inference

- We use Fisher approach, focused on deriving exact p-values for sharp null hypotheses regarding the effect of treatments.
- Under a sharp null hypothesis all potential outcomes are known from the observed values of the potential outcomes.
- The Fisher Exact P-values approach entails three steps: (i) the choice of a sharp null hypothesis, (ii) the choice of test statistic, and (iii) the measure of extremeness (p-values).
- We focus on the sharp null hypothesis of no effect of the treatment for any unit (class) in the population:

$$H_0\text{: } Y_k(0) = Y_k(1) \text{ for all } k$$

- Under H_0 $Y_k(0) = Y_k(1)$ for all k.
- Test statistics:

$$S_{ave} = \text{Difference in average outcomes by treatment status}$$

$$S_{rank} = \text{Difference in average ranks for treated and control units}$$

IRPET

Cluster randomization inference: results

Observed values of the test statistics and p-values for H_0: $Y_k(0) = Y_k(1)\ \forall k$ *versus* H_1: $\exists k$: $Y_k(0) \neq Y_k(1)$

Outcome variables	S_{ave}	*p*-value	S_{rank}	*p*-value
Percentage of failures	-2.78	0.6698	-1.78	0.7104
Percentage of postponements of the evaluation	5.87	0.2320	5.77	0.2270
Percentage of drop-outs	-2.41	0.7734	-3.12	0.7744
Absence rate (%)	-0.15	0.9434	-1.79	0.6804
Percentage of failures + Drop-out	-5.19	0.4554	-5.25	0.4794

Table shows the observed values of the test statistics and the *p*-values against the alternative that, at least for some units, there is a non-zero effect.

<u>The test statistics show some evidence that the new teaching method reduces the percentage of drop-outs and failures and the absence rate, and increases the percentage of postponements of the evaluation, but the results do not prove statistically significant.</u>

The multilevel regression model

- Individual-level analysis based on multilevel models allow us to adjust for both individual-level and cluster-level characteristics and to easily obtain estimates of intraclass correlation.
- We consider generalized linear mixed models with probit link for binary outcome variables and linear mixed models for continuous outcomes.
- For each outcome variable we include individual-level and cluster-level variables and group-averages of the first level variables as explanatory factors.
- Group-averages of the first level variables allow us to account for the presence of correlation between individual-level variables and cluster effects, as well as for the presence of interference between students belonging to the same class.

Results: Individual-level Analysis

Outcome variable	W_{ki}	Variance: Cluster-level (Residual)	$E[Y_{ki}(0)]$	$E[Y_{ki}(1)]$	$E[Y_{ki}(1)]-E[Y_{ki}(0)]$
Failures	-0.258 (0.039)	0.000	0.157	0.105	-0.052
Postponement of the evaluation	0.245 (0.013)	0.000	0.253	0.337	0.084
Drop-out	-0.047 (0.799)	0.000	0.015	0.014	-0.002
Absence rate (%)	-1.024 (0.883)	4.284 (52.764)	14.835	13.810	-1.024
Failure + Drop-out	-0.217 (0.111)	0.000	0.223	0.164	-0.059

The model provides statistical significance that the new teaching method reduces failure rate and increases the probability of postponement of the evaluation.

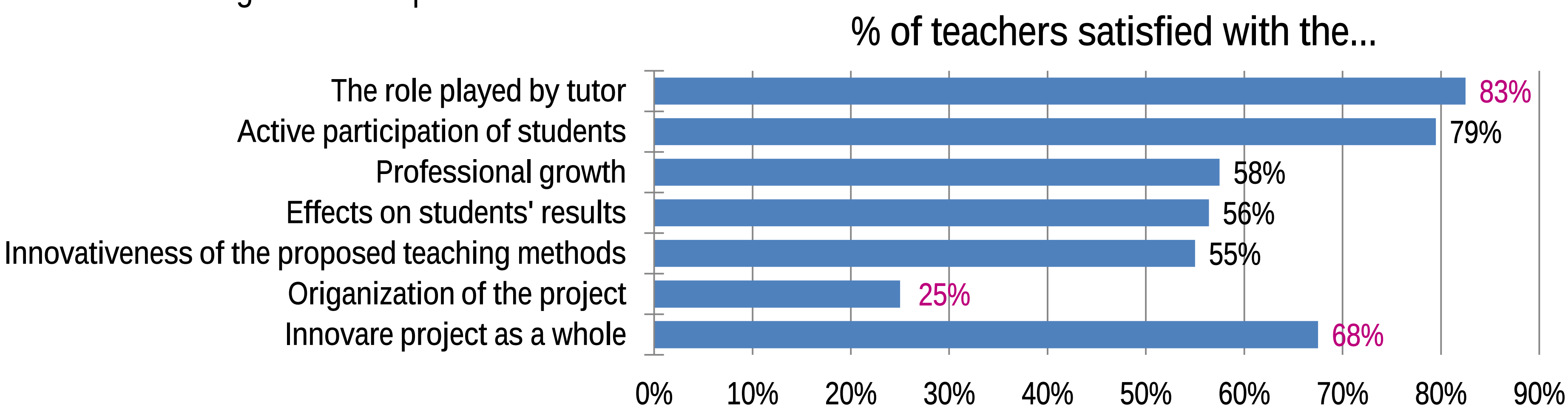
Qualitative approach: the Focus Group and the questionnaire

The goal of the focus groups is twofold:

1. to record the point of view of the "trainers" (tutors) and “trainees" (teachers);
2. **discuss the main objectives** of the project in relation to the **results achieved**, **problems** encountered, the **response of the students**, the future prospects.

Main results:

1. The Focus revealed a **general satisfaction with the project Innovare**, practically **unanimous among the tutors**, less pronounced but still a **majority even among teachers.**
2. Particularly, **teachers stressed the effect of re-motivation descending from the relationship between trainers and trainees**, which receives the highest level of satisfaction expressed by teachers answering to a short questionnaire.



Conclusion

1. Two types of quantitative analysis were carried out to assess the impact of the project INNOVARE on drop-out: a cluster-level analysis and an individual-level analysis using the potential outcome approach to causal inference.
2. Both methods show a slight decrease in the probability to fail, to drop-out, and in the absence rate, and conversely an increase in the probability of postponement of the evaluation, linked to participation in INNOVARE.
3. These effects, however, appear to be quantitatively modest and statistically not reach significance, due to insufficient sample size; using the multilevel regression model statistical significance is found for some outcome variables.
4. These results are promising when considered together with those emerged from the Focus Group analysis: notwithstanding organizational problems, the program managed to re-motivate teachers.

Istituto Regionale Programmazione Economica della Toscana

International Workshop of Applied Economics of Education

The effects of a dropout prevention program on secondary students' outcomes

Enrico Conti, Silvia Duranti, Alessandra Mattei, Fabrizia Mealli, Nicola Sciclone

Catanzaro, 22-24 giugno 2014

Advantages and drawbacks of both methods

Advantages of Cluster Level approach

1. In a cluster-level approach, we can always conduct exact statistical inferences without introducing parametric assumptions and we can easily adjust for background characteristics.
2. Also, a cluster-level analysis is correct and valid irrespective of the strength of the intraclass correlation, because it implicitly accounts for all sources of variability.
3. Results coming out of this analysis are exact and valid irrespective of the number of group assigned to each treatment status (e.g., Small et al., 2008, Imbens and Woolbridge, 2009; Mealli et al., 2011).

Advantages and Drawbacks of individual level approach

1. An individual-level analysis which accounts for the presence of intraclass correlation, using e.g., mixed effect regression models, may lead **to more powerful model-based tests if the model is well specified** than group randomization inference (Braun and Feng, 2001). But in the literature, widely accepted guidelines for the numbers of cluster required to ensure validity of statistical inferences. Say that results from studies enrolling fewer than 20 clusters per intervention group should be interpreted with caution (Duncan et al., 1998). The INNOVARE study involves only 18 treated classes and 35 control classes.
2. When applicable, multilevel models offer several advantages over cluster-level analyses. Specifically, multilevel models allow one to
   - obtain estimates of intraclass correlation more naturally, which can be used to design future studies;
   - adjust for background covariates at both individual- and cluster-level;
   - investigate sources of heterogeneity in the treatment effect, including interaction between the treatment variable and some specific covariates;
   - extend the analyses to more complex data structures more easily, involving more than two levels.

The cluster level analysis

In the INNOVARE study, we implement cluster-level randomization inference adjusting for differences in the observed cluster-level covariates using subclassification on the propensity score (Rosenbaum and Rubin, 1983).

There exist some differences in background pretreatment variables between the treatment group and the control group.
In order to account for these differences in the observed pretreatment variables we use subclassifications on propensity score.

The conditional probability of receiving a treatment given pretreatment characteristics under the assumption that the treatment is strongly ignorable: $\Pr(W_k=1 \mid Y_k(0), Y_k(1), X_k)=\Pr(W_k=1 \mid X_k)$, and $0 < \Pr(W_k=1 \mid X_k) < 1$, $k=1, \ldots, K$ (Rosenbaum and Rubin, 1983).

Strong ignorability amounts to assuming that within cells defined by the values of pre-treatment variables, the treatment is randomly assigned. Under this assumption we can view INNOVARE as a stratified cluster randomized experiment

(Small et al., 2008). Rosenbaum and Rubin (1983) show that if the exposure to treatment is random within cells defined by the covariates, it is also random within cells defined by propensity score: $\Pr(W_k=1 \mid Y_k(0), Y_k(1), e(X_k))=\Pr(W_k=1 \mid e(X_k))$, where $e(X_k)=\Pr(W_k=1 \mid X_k)$, is the propensity score for the *kth* class, $k=1,\ldots, K$.

1. In our analysis, the propensity score is estimated using a logit regression model.

2. Based on the estimated propensity score, we restrict the analysis to the subsample of classes that satisfies an overlap or common support condition. Specifically, we discard four control classes with propensity score values lower than the minimum propensity score value for the treated classes, and one treated class with propensity score greater than 0.9, which is an extremely high value in our sample.

3. Then, we re-estimated the propensity score using the selected subsample of 48 classes and use it for adjusting treatment comparisons for differences in background covariates using subclassification.

Results

The observed values of the test statistics and the *p*-values against the alternative that, at least for some units, there is a non-zero effect : H_1: \$ *k*: $Y_k(0)$ ¹ $Y_k(1)$. The *p*-values are estimated using 10,000 draws from the randomization distribution.

<u>The test statistics show some evidence that the new teaching method reduces the percentage of drop-outs and failures and the absence rate, and increases the percentage of postponements of the evaluation.</u>

Table 3. Observed values of the test statistics and p-values for the sharp null hypothesis H_0: $Y_k(0) = Y_k(1)$ " *k* against the alternative H_1: \$ *k*: $Y_k(0)$ ¹ $Y_k(1)$

Outcome variables	S_{ave}	*p*-value	S_{rank}	*p*-value
Percentage of failures	-2.78	0.6698	-1.78	0.7104
Percentage of postponements of the evaluation	5.87	0.2320	5.77	0.2270
Percentage of drop-outs	-2.41	0.7734	-3.12	0.7744
Absence rate (%)	-0.15	0.9434	-1.79	0.6804
Percentage of failures + Drop-out	-5.19	0.4554	-5.25	0.4794

Conclusioni

Outcome Variables	Observed Mean Differences	S_{ave}	*p-value*	S_{rank}	*p-value*
Percentage of failures	-2.96	-2.78	0.6698	-1.78	0.7104
Percentage of postponements of the evaluation	2.93	5.87	0.2320	5.77	0.2270
Percentage of drop-outs	-0.21	-2.41	0.7734	-3.12	0.7744
Absence rate (%)	1.08	-0.15	0.9434	-1.79	0.6804
Percentage of failures + Drop-out	-3.18	-5.19	0.4554	-5.25	0.4794

teachers ?	Enough / very much	83%	77%	81%
How much are you satisfied with the role played by tutor/experts??	Not at all / not very	18%	17%	23%
	Enough / very much	83%	80%	78%
Have students participated actively in the INNOVARE project?	Not at all / not very	53%	41%	36%
	Enough / very much	47%	55%	64%

Evaluation design: the quantitative approach

The quantitative statistical approach verifies the presence or absence of a causal link between the treatment of the classes involved in the project INNOVARE and some outcome variables.
We adopted both a cluster-level analyses using the potential outcome approach to causal inference (e.g., Rubin 1974, 1978, 1990a,b, 2005) and an individual-level analyses using a multilevel regression model.

1. A cluster-level analysis may provide useful information on the effectiveness of the intervention in reducing high-school drop-out. Here, drop-out is viewed as a social problem and focus is on interventions that can limit school drop-out as a whole. **The class is the natural unit of inference and standard methods for the analyses of randomized experiments** can be applied at the cluster level.

2. Multilevel regression analysis. An individual-level analysis aims at assessing whether the innovative teaching method has a causal effect on **the student probability of dropping-out of school**. In this case, the unit of assignment (class) is different from the unit of analysis (student), and the lack of independence among student in the same class, i.e., **the presence of intraclass correlation, creates special methodological challenge** and cannot be ignored.