METHODOLOGY REPORT

Randomization Process in Assessing the Impact of the Program Jovem de Futuro

Authors: Ricardo Paes Barros & Samuel Franco

INSTITUTO UNIBANCO
Methodology Report: Randomization Process in Assessing the Impact of the Program Jovem de Futuro
Authors: Ricardo Paes Barros // Samuel Franco
ISBN 978-85-60154-04-3

Resumo

O presente relatório tem como objetivo explicitar as escolhas metodológicas e processuais adotadas para aleatorização das escolas do Jovem de Futuro e o método de avaliação empregado para estimação do impacto do programa.

Após o término da segunda geração (executada entre 2011 e 2014), o Jovem de Futuro passou por uma reestruturação em suas frentes de trabalho e isso levou ao desenvolvimento de uma terceira geração, implementada nos estados do Espírito Santo, Piauí e Pará entre 2015 e 2017. As mudanças propostas acarretaram na necessidade de pactuação de um método que viabilizasse a realização de avaliações

Abstract

This report aims to explain the choices of methods and processes used to randomize the Jovem de Futuro schools, as well as the assessment method adopted to estimate the program's impact.

After the end of the second generation (carried out between 2011 e 2014), the work fronts of the Jovem de Futuro were restructured, leading to the development of a third generation, implemented in the states of Espírito Santo, Piauí and Pará from 2015 to 2017. The proposed changes raised the need to agree on a method to carry out impact assessment in the three states comprised in the third generation.

de impacto nos três estados abrangidos pela terceira geração. A opção por um método experimental de seleção via aleatorização das escolas e diferenciação do tempo de entrada no programa criou bases sólidas para um processo de avaliação metodologicamente estruturado e passível de validação. Deste modo, as escolas selecionadas para receber o benefício no início do Programa (1º ano) foram denominadas escolas de tratamento e aquelas selecionadas para o último ano de implementação (3º ano), escolas de controle. PALAVRAS-CHAVE: Aleatorização; Avaliação Experimental; Grupos de tratamento e controle; Jovem de Futuro; Gestão Escolar; Ensino Médio	The option for an experimental selection method that randomizes the schools and distinguishes when they joined the program created solid bases for a methodologically structured assessment process amenable to validation. Thus, the schools selected to receive the benefit at the beginning of the program (first year) were called "treatment schools", and those selected for the last year of implementation (third year), "control schools". KEYWORDS: Randomization; Experimental assessment; Control and treatment groups; Jovem de Futuro; School management; High School.

SUMMARY

1. Original Design for School Selection

The choice of method

Program Jovem de Futuro was developed bearing in mind that "efficient, participative management focused on positive learning results can decisively influence the quality of education offered at school" (UNIBANCO, 2012).

In 2007, Instituto Unibanco deployed the Program in a group of public secondary schools in Belo Horizonte and Porto Alegre. In 2010, the Program was expanded to Rio de Janeiro, São Paulo and the of the valley of the Paraíba river. The implementation of the Program in these five regions, with financial resources and technology from Instituto Unibanco, became known as the pilot. In the ensuing years, the pilot was the cornerstone of a partnership with the Ministry of Education (MEC) to merge the Program Jovem de Futuro with PRoEMI, the Innovative High School Program. This dissemination became know as the Program's 2nd Generation, when it was widely implemented in the states of Ceará, Goiás, Mato Grosso do Sul, Pará and Piauí with resources from the Federal Government, technology from Instituto Unibanco and partnership with the State Departments of Education of each state.

During the 2nd Generation (2011-2014), not only the Program Jovem de Futuro (PJF) itself, but also its assessment and the randomization process of participating schools were restructured as the Program's impact in those regions was reviewed. This restructuring, which became known as the 3rd Generation, was carried out in the state of Espírito Santo, where intervention began in 2015, and was replicated in the 2015-2017 Phase in the states of Piauí and Pará. The main elements of the restructuring are highlight below:

- The *Program Jovem de Futuro*, now detached from the ProEMI, sets goals for schools, with special attention to vulnerable schools and to a follow-up system;
- *Impact Assessment* in the 12th grade (senior high school year) of the schools under the Program's intervention will take place over two years, not three as before;

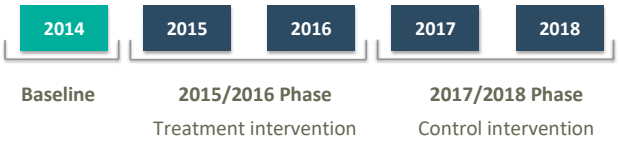
The Program's implementation required that a method be chosen to ensure the proper assessment of the selected schools and the measurement of the Program's impact (experimental method).

Thus, an experimental, randomization-based selection method was adopted,[1] which allowed the Program to be implemented on the solid foundations of a methodologically structured selection process, amenable to validation, to choose the participating schools. An essential mechanism for assessing the impact of the Program's results, was also included, because the selection process makes it possible to classify the treatment and control schools.

Treatment and control groups

The schools selected to receive the benefit at the onset of the Program (first year) were called *treatment* schools, while those selected for the last year of treatment (third year) were called *control* schools, because they will receive the benefit only after the intervention in the treatment schools is finished. Because of this structure, the intervention is organized as follows:

[1] "*Randomization*, in which the selection into the treatment and control groups is random within some well-defined set of people. In this case there should be no difference (in expectation) between the two groups besides the fact that the treatment group had access to the program […]." (BAKER, 2000)



Homogeneity of the groups

One must consider the possibility that, by randomly choosing among the schools enrolled in the Program, without replacement, a rare event may occur in which a certain outstanding characteristic will be more present in one of the two groups; for instance, the group of schools selected for treatment might have performed better in previous assessments than the control group, or vice-versa.

To avoid this problem, a viable procedure is to form groups of similar schools and then draw lots within each group. This ensures that among similar schools, one will be in the treatment group and the other in the control group, thus maintaining a "balance" of characteristics between the schools that will be compared.

Randomization should, therefore, generate optimal control and treatment groups in terms of capturing the Program's impact. Our working hypothesis is that *the groups deemed ideal are those whose performance evolution remains the same in the absence of the Program*.

Performance evolution is controlled by two traits: (i) the schools' average grade in Portuguese Language and Mathematics tests at the Program's baseline (to ensure that schools with the same evolution but very different starting grades are not part of the same group); and (ii) the Socioeconomic Level of the School (NSE), an index that measures the vulnerability of a school with respect to the redesign of the Program.

Tests to define the best method

We postulated that groups deemed ideal are those whose performance evolution remains the same in the absence of the Program.

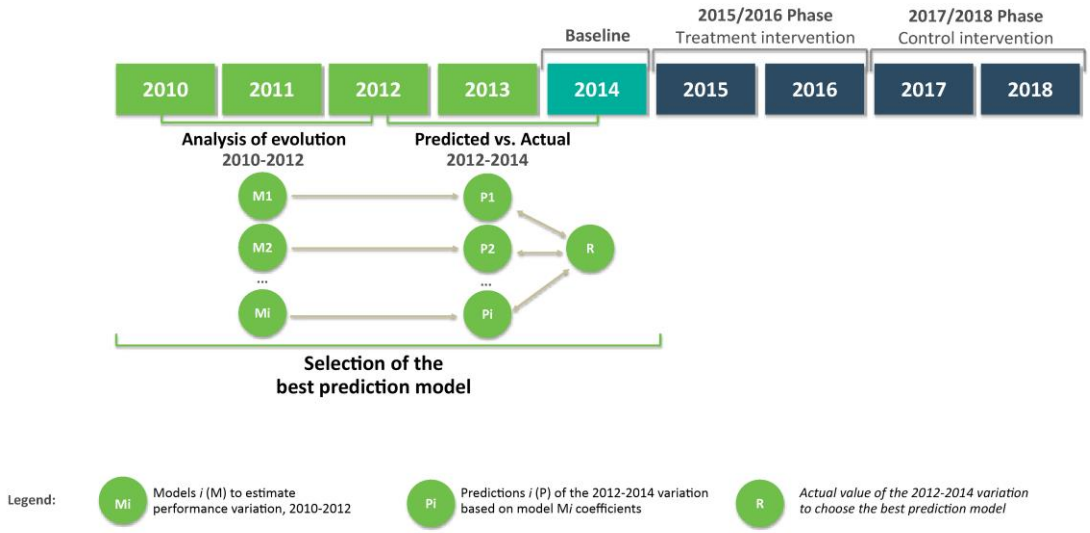
- *Problem: The evolution parity of our hypothesis refers to the 2015/2016 Phase of the Program*, that is, to a future moment;
- *Solution: To estimate the evolution of the groups for 2015/2016 without the Program's intervention.*

To estimate the future evolution of these groups, we must resort to past information about the performance of the schools. Because the Program's baseline is 2014, and 2015 and 2016 are the years of the first phase, we will use the 2010-2012[2] evolution to predict the 2012-2014 evolution, seeing we have information available for both periods. At the end, we will compare actual evolution in 2012-2014 with the evolution predicted by the 2010-2012 model. The method chosen for the pairing of schools will be the one that ensures the best prediction.

Once the best prediction model for each region is chosen, the process of randomizing, pairing and selecting schools can begin.

[2] The test's original design considers the 2010-2012 evolution to make the 2012-2014 prediction. However, the actual range used will vary depending on the availability of data in each region.

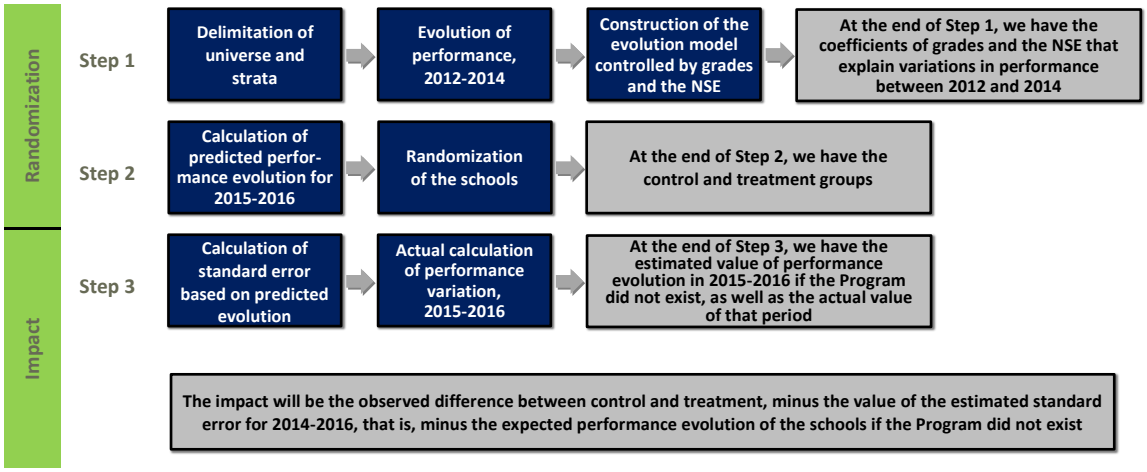
THEORETICAL FRAMEWORK OF THE PREDICTION MODEL: THE TEST STRUCTURE



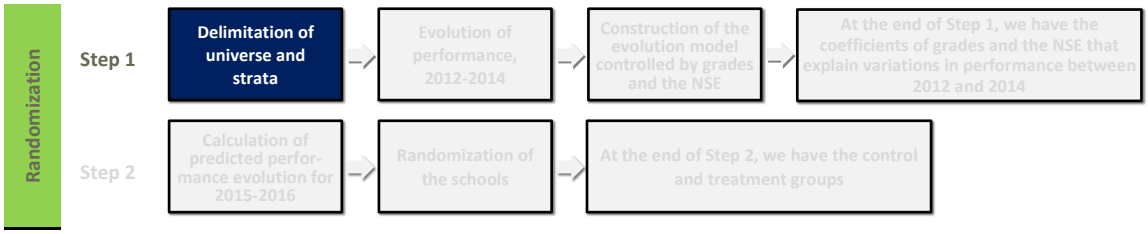
Randomization and pairing

In keeping with the design of the Program, the experimental nature of the assessment and the need for homogeneity in the control and treatment groups, randomization was carried out in two steps. The third step is estimating the impact, as shown in the diagram below:

OVERVIEW OF THE PROCESS



Step 1: Delimitation of universe and strata



Step 1 involved delimiting the universe and strata of schools that would take part in the 2015-2016 Phase of the Program and in the assessment of impact at the end of 2016. The universe and stratification of schools were defined according to their management capacity and the priorities of each Department of Education. Therefore, each of the Program's locales of intervention has a distinct peculiarity. The only counterpart required by Instituto Unibanco was for the Department to provide support for a minimum number of schools not participating in the 2015-2016 Phase in order to make it possible to assess the impact at the end of 2016. The group not participating in the 2015-2016 Phase will join the Program in the two following years (2017-2018 Phase).

The strata were organized into two levels. The first takes into account the territorial characteristics of the state (the number of municipalities, regional offices and so on) pertaining to the realities of each Department. The second level, after territorial delimitation, is the stratification according to vulnerability of the schools, as follows:

Level 2 stratification: according to the school's vulnerability:

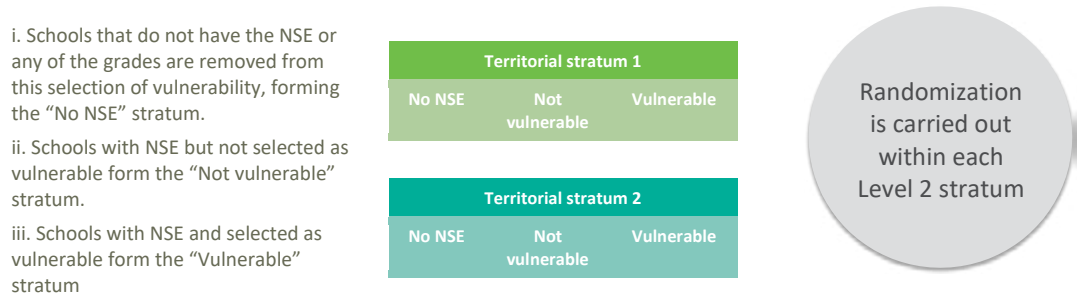
i. The Department and Instituto Unibanco determined the number of vulnerable schools they would be able to assist during the Program's intervention;
ii. For each school, their percentile was calculated for three different indicators: the Socioeconomic Level of the School (NSE), the average grade in Portuguese Language and Mathematics in 2014

(Program baseline) and the variation in the average grade in those two subjects from 2012 to 2014 (recent observed performance evolution);

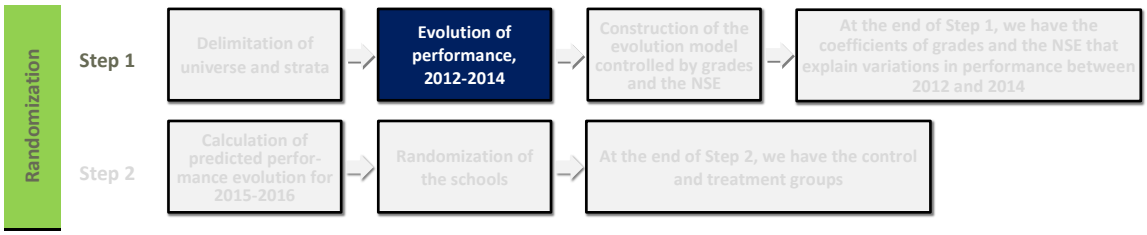
iii. The maximum value of the school's percentile of the three indicators was calculated;

iv. Within each territorial stratum, the schools were ranked according to the lowest value observed in the percentiles of the three indicators; and

v. Schools were selected until the number of vulnerable schools determined by each Department was reached.



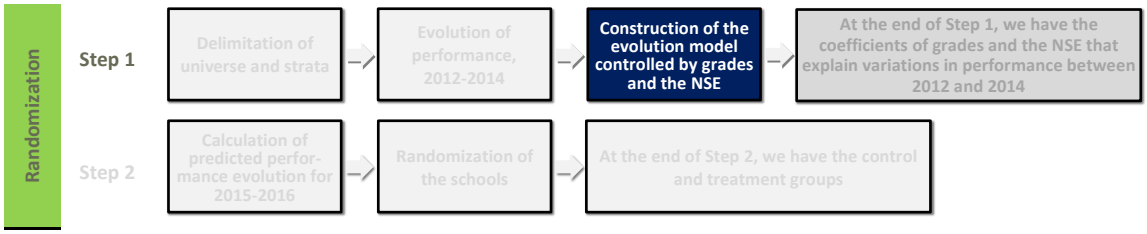
Step 1: Evolution of performance, 2012-2014



Once the universe was chosen, we calculated the school's performance in state-wide assessments over a period of at least two years prior to the implementation of Phase 1 of the Program. Seeing that Phase 1 would begin in 2015, the evolution of performance evolution had to be considered for at least 2012-2014.

Based on performance variation between 2012 and 2014, it was possible to determine the similarity between the control and treatment schools, ensuring the homogeneity of both groups. This was necessary because we started from the hypothesis that similar schools are those with similar evolution in school performance.

Step 1: Construction of the evolution model controlled by grades and the NSE



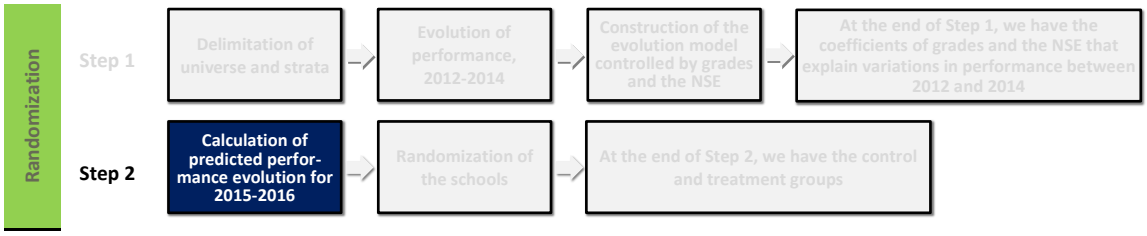
To ensure that schools with similar performance evolution were actually similar, the analysis of their evolution was controlled by three variables: (i) the grades obtained in Portuguese Language and Mathematics tests in 2012, thus making sure the schools started from a similar performance level, and (ii) the socioeconomic level of the school as measured by the NSE index, making sure that the socioeconomic realities of the schools were similar.[3]

Thus, schools that evolved similarly between 2012 and 2014 also performed similarly in 2012 and had similar socioeconomic characteristics. Accordingly, we built a model of evolution controlled by the grades obtained in Portuguese Language and Mathematics tests and by the NSE. The variation of the average performance in the tests between 2012 and 2014 (dependent variable) was determined by the grade in Portuguese Language in 2012 (initial year), by the grade in Mathematics in 2012, and by the NSE (independent

[3] In calculating the coefficients of the model, schools without grades remain part of the randomization process and take on the estimated value of the coefficients of the other schools.

variables). Therefore, the coefficients of this regression correspond to how much of the performance variation between 2012 and 2014 can be explained by the independent variables.

Step 2: Calculation of predicted performance evolution for 2015-2016



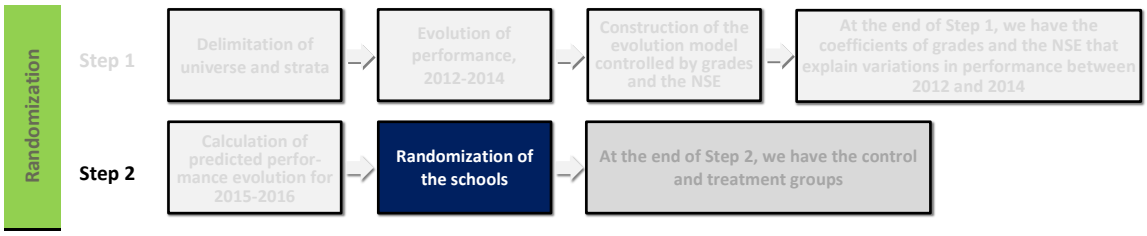
Using the coefficients of the 2012-2014 model, we assumed the behavior of future evolution (2015-2016) would be similar to that observed in 2012-2014. Thus, based on these coefficients, we estimated the *evolution of predicted performance for 2015 and 2016* for each school if the Program didn't exist. This estimate is controlled by the grades in Portuguese Language and Mathematics test of the Program's baseline (2014) and by the schools' NSE.

Thus, each school in the universe has its own expected (predicted) performance evolution for the 2015-2016 Phase absent the Program's intervention. In the tests that were carried out, prediction model was chosen as the best one to forecast the future variation of a school's results in the absence of the Program.

With the predicted value of performance variation in 2016, we determined the evolution that schools would have in the absence of the Program. By estimating a standard error based on this value, we can, at the end of the intervention, isolate the Program's impact, which will be the difference in performance of the control and treatment groups in 2015-2016. The Program's impact becomes identifiable when the evolution is greater than the standard error.

Schools without information on 2014 grades (baseline) will not be part of the assessment.

Step 2: Randomization of the schools

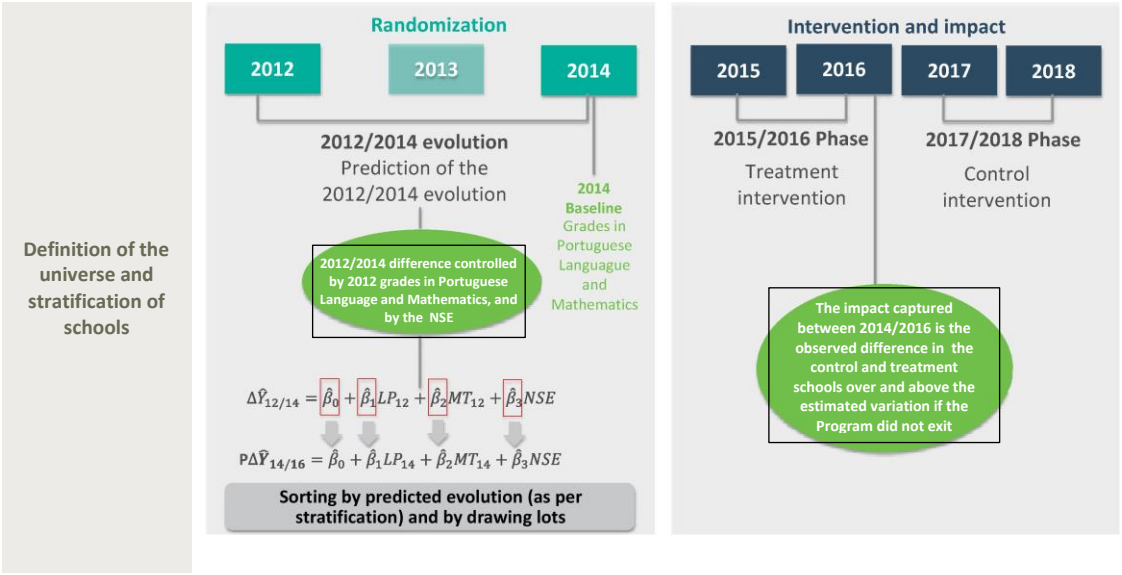


The predicted performance evolution is used to rank schools within each stratum. The number of groups formed within each stratum was determined by the number of schools where the Department of Education was not willing to participate in the Program's intervention in the 2015-2016 Phase.

Randomization was then performed. A control school was retained for each group of schools and the treatment schools were distributed evenly between the groups.

For instance, if the universe consisted of 100 schools and the Education Department determined that only 20 would be control schools, each group of schools would comprise four treatment schools and one control school (for a total of 20 groups). If the universe consisted of 90 schools and only 25 were control, the groups could be formed in two ways: either two treatment schools and one control school (10 groups) or three treatment schools and one control school (15 groups).

SUMMARY CHART



Validating the randomization

Randomization validation was based on T-tests for equal means and on the Kolmogorov-Smirnov test. For the performance indicators of the treatment and control schools, the observed difference must be statistically non-significant for the treatment and control groups to be considered equal with a significance level of 5%. If the results of these tests differ, the analysis is deemed inconclusive.

T-Test for Equal Means

We considered dij the observed difference in student performance between schools i and j. Assuming, as a null hypothesis, the difference between the means is zero ($Ho: m_d = 0$) and, therefore, as an alternative hypothesis, the existence of a difference between the means ($Ho: m_d \neq 0$), the T-test is given by:

$$T = \frac{\bar{d} - m_d}{\frac{s_d}{\sqrt{n}}}$$

Where $\bar{d}$ is the mean of the differences between the pairs, S_d is the standard deviation estimated based on the sampling variance of the differences, and total paired observations have a significance level of 10% (90% confidence). By accepting the null hypothesis, we concluded that the pairs are statistically equal, that is, the difference between the means can be considered null.

Kolmogorov-Smirnov (KS) Test

This test is based on the maximum vertical distance between the cumulative frequencies of the distribution function of each independent random sample of sizes n1 and n2, and distribution functions Y(X1) and Z(X2). The test checks whether or not there is a statistical difference between the variables being studied based on two hypotheses: H0, there is no difference between the samples [Y(X1) = Z(X2)]; and H1, there is a difference between the samples [Y(X1) ≠ Z(X2)].

As the test concerns the maximum vertical distance [KS=max / Y(X1) – Z(X2)], it shows whether the greatest observable distance between schools is statistically significant, allowing them to be considered different, compared to the tabulated ks value or the p-value of the statistic. If we accept the null hypothesis, we conclude that the greatest observed difference can be considered null and, therefore, the samples can be considered statistically equal.

2. Randomization in the State of Piauí

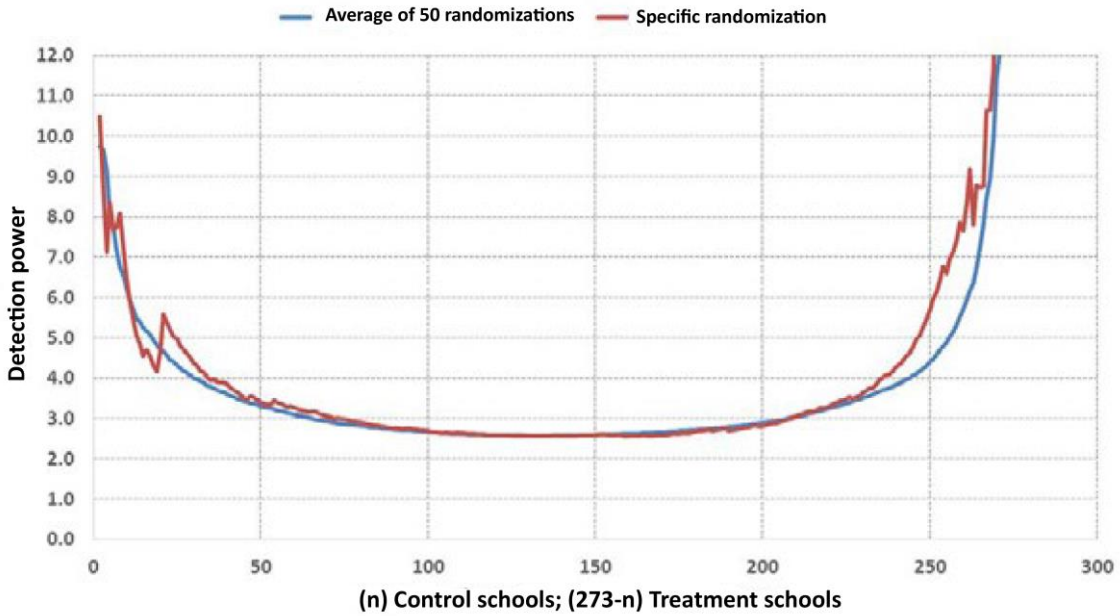
Tests

The method of pairing schools by predicting their performance evolution from 2014 to 2016 was selected for the process of randomly choosing schools in Piauí. For the test, a group of 273 schools was considered and the following models were tested: (i) the prediction model, controlled by grades and the NSE of the schools, to pair up and select the treatment and control groups; and (ii) the random model to select the treatment and control groups.

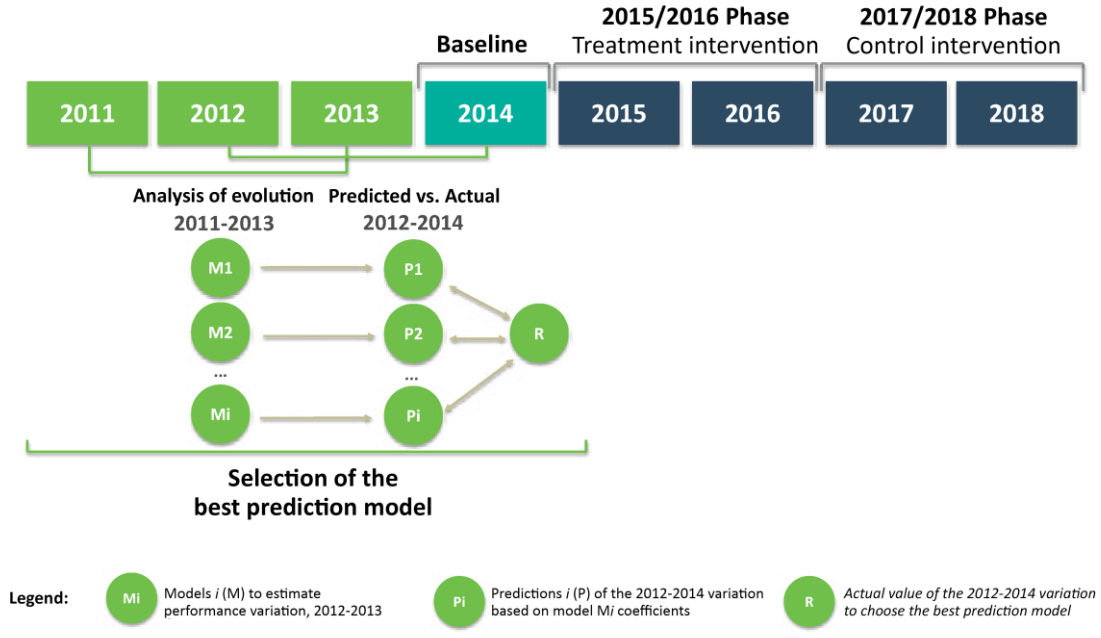
The criterion used to choose the best model was based on a lower impact detection power, that is, the chosen model would be the one most sensitive to capture the Program's impact. Thus, any future difference found between control and treatment groups over and above the value of the detection power can be attributed to the Program's impact.

The prediction model (i) considered 123 control schools and 150 treatment schools, and its detection power was 2.8. The random model (ii) considered the same number of control and treatment schools, and its detection power was 2.6. Given the similarity of results, the prediction method was chosen for the pairing, so as to use the same methodology that was the best fit in other regions. Thus, the model chosen to perform the randomization process was (i), the prediction model controlled by grades and NSE of the schools.

DETECTION POWER WITH A BALANCED NUMBER OF CONTROL AND TREATMENT SCHOOLS: PIAUÍ, 273 SCHOOLS – RANDOM PAIRING



THEORETICAL FRAMEWORK OF THE PREDICTION MODEL: THE TEST STRUCTURE, PIAUÍ



It should be noted that, in Piauí, the original design of the prediction test model had to be adjusted for the 2011-2013 period in order to test calculate the predictions for 2012-2014.

The overlapping of test intervals was necessary because of the inexistence of data for 2010 data and to maintain the model's 2-year evolution design.

Delimitation of universes and strata

At this stage of redesigning the Program Jovem de Futuro (Phases 2015-2016 and 2017-2018), the agreement between Instituto Unibanco and the State Department of Education resulted in a group of 275 schools eligible to participate in the Program.

The criteria used by the Department to delimit the universe were, in this order:

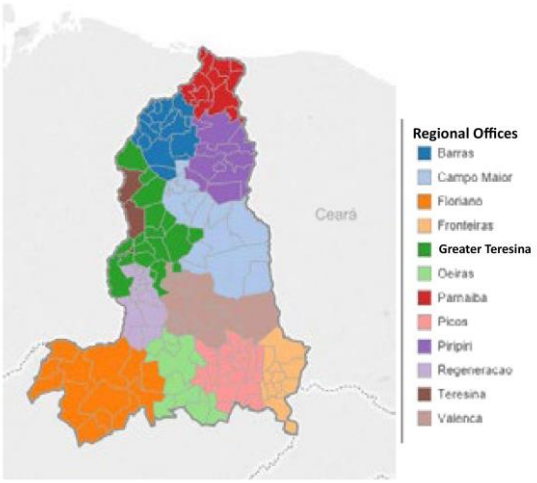
(i) To belong to any of the 15 chosen Regional Offices (according to the Education Department's capacity to act);

(ii) To have more than 100 students enrolled in regular high school in 2015.

Table – Universe of schools and Domain: Piauí, Assessment by JF

Selection	Number of schools	Number of groups			Domain		Vulnerable	
		Total	Pairs	Trios	T	C	Pairs	Schools
Total no. of schools on file	468	–	–	–	–	–	–	–
Among the 15 selected Regional Offices	371	–	–	–	–	–	–	–
With at least 100 high school students in 2015	275	134	131	3	137	134	10	20
Teresina	73	35	33	2	37	35	0	0
North	60	29	29	0	29	29	2	4
Center	65	32	32	0	32	32	4	8
South	77	38	37	1	39	38	4	8

Regional Office		Region
1	PARNAÍBA	NORTH
2	BARRAS	NORTH
3	PIRIPIRI	NORTH
5	CAMPO MAIOR	CENTER
7	VALENÇA	CENTER
18	GREATER TERESINA	CENTER
6	REGENERAÇÃO	SOUTH
8	OEIRAS	SOUTH
9	PICOS	SOUTH
10	FLORIANO	SOUTH
16	FRONTEIRAS	SOUTH
4	TERESINA (CENTER NORTH)	TERESINA
19	TERESINA (SOUTH)	TERESINA
20	TERESINA (NORTHEAST)	TERESINA
21	TERESINA (SOUTHEAST)	TERESINA



Initially, these 275 schools were stratified into four groups, each comprising a number Regional Offices of education.

Of the 275 schools selected for the universe, four lacked evolution performance data for 2012-2014. As a result, the baseline could not be defined and, therefore, they were excluded from the assessment, reducing the number of assessed schools to 271. The schools excluded from the assessment are listed below. They remained in the randomization only to define the Phase in which they will join the Program.

- 22017011 – U.E. Presidente Castelo Branco (NORTH);
- 22001557 – U.E. Honorina Tito (NORTH);
- 22017992 – U.E. Afonso Mafrense (CENTER);
- 22028021 – U.E. Cristino Castelo Branco (TERESINA).

Table – Level 1 stratification – territorial criterion: Piauí

Region	Schools			Groups	
	Total	Control	Treatment	Pairs	Trios
Teresina	72	35	37	33	2
North	58	29	29	29	0
Center	64	32	32	32	0
South	77	38	39	37	1
Total	**271**	**134**	**137**	**131**	**3**

For the second level of stratification (specifying vulnerable schools), the Department decided that the domain would be limited to 20 schools.

- Teresina – no vulnerable stratum;
- North – 2 pairs, 4 schools;
- Center – 4 pairs, 8 schools;
- South – 4 pairs, 8 schools.

Table – Level 2 stratification – vulnerability criterion: Piauí

School code	School	Region	Regional Office(GRE)
22135383	UNIDADE ESCOLAR PEDRO MENDES PESSOA	CENTER	GREATER TERESINA
22029869	UNIDADE ESCOLAR PROFA ELISA SOUSA	CENTER	GREATER TERESINA
22132740	UNIDADE ESCOLAR MENINO JOÃO PEDRO	CENTER	GREATER TERESINA
22030158	UNIDADE ESCOLAR FILINTO REGO	CENTER	GREATER TERESINA
22019294	UNIDADE ESCOLAR RAIMUNDO MARTINS	CENTER	GREATER TERESINA
22048456	UNIDADE ESCOLAR JOÃO DE DEUS CARVALHO	CENTER	VALENÇA
22135901	UNIDADE ESCOLAR CEZAR LEAL	CENTER	GREATER TERESINA
22034072	UNIDADE ESCOLAR BRIOLANJA OLIVEIRA	CENTER	CAMPO MAIOR
22011170	UNIDADE ESCOLAR SÃO JOSÉ	NORTH	BARRAS
22126880	UNIDADE ESCOLAR SANTA TERESINHA	NORTH	BARRAS
22007431	UNIDADE ESCOLAR JOSÉ AMAVEL	NORTH	BARRAS
22137556	UNIDADE ESCOLAR JOÃO ODORICO	NORTH	BARRAS
22083480	UNIDADE ESCOLAR DOM EDILBERTO DINKELBORG	SOUTH	OEIRAS
22135030	CENTRO EDUCACIONAL SEBASTIÃO DE SOUSA	SOUTH	FRONTEIRAS
22136509	UNIDADE ESCOLAR JOÃO ANTONIO DA VERA	SOUTH	PICOS
22134484	UNIDADE ESCOLAR HELVIDIO NUNES	SOUTH	PICOS
22086528	UNIDADE ESCOLAR NOSSA SENHORA DO PATROCÍNIO	SOUTH	FRONTEIRAS
22060596	UNIDADE ESCOLAR JOSÉ SALUSTIANO DA SILVA	SOUTH	FLORIANO
22083065	UNIDADE ESCOLAR CLEMENTINO MARTINS	SOUTH	PICOS
22083189	UNIDADE ESCOLAR JOAQUIM BORGES DE OLIVEIRA	SOUTH	PICOS

The final delimitation of the universe of schools for the randomization of the control and treatment groups included 271 schools divided into 134 groups:

Table – Level 1 stratification – Territory: Piauí

Region	Schools			Groups		
	Total	Control	Treatment	Non-vulnerable pairs	Non-vulnerable trios	Vulnerable pairs
Teresina	72	35	37	33	2	0
North	58	29	29	27	0	2
Center	64	32	32	28	0	4
South	77	38	39	33	1	4
Total	**271**	**134**	**137**	**121**	**3**	**10**

Of the 271 schools taking part in the randomization process for selection of the control and treatment groups, 232 were paired according to the prediction of performance variation for the 2014-2016 period; 20 were paired according to their vulnerability; and 19 were randomly paired because they were not vulnerable and lacked sufficient information for the 2014-2016 prediction.

The 2012-2014 model for estimating coefficients for the predictor of performance evolution between 2014 and 2016 generated the following statistics. Once the predicted value was calculated, schools were sorted within their strata and randomized.

Summary of results

Regression statistics	
Multiple R	0.332
R squared	0.110
Adjusted R squared	0.099
Standard error	12.127
Observations	246

ANOVA

	GL	*SS*	*MS*	*F-value*	*Significant F*
Regression	3	4417.952	1472.651	10.014	0.000
Residuals	242	35589.096	147.062		
Total	245	40007.048			

	Coefficients	Standard error	Stat T	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intersection	57.10134	12.76004	4.47501	0.00001	31.96641	82.23627	31.96641	82.23627
NSE	-1.57311	3.57610	-0.43990	0.66040	-8.61736	5.47114	-8.61736	5.47114
LP 2012	-0.32077	0.09797	-3.27413	0.00121	-0.51375	-0.12778	-0.51375	-0.12778
MT 2012	0.08149	0.10106	0.80627	0.42088	-0.11759	0.28056	-0.11759	0.28056

Validation

Tables with the values of the T-test for equal means and the Kolmogorov-Smirnov test are shown below.

Table – Mean and Kolmogorov tests for the Mathematics baseline results in the Treatment and Control schools of the Program Jovem de Futuro: Piauí, 2014

Indicators	Mean (%)		Difference	P-value of T-test of equal means	Statistics of Kolmogorov test	P-value of Kolmogorov T-test
	Treatment	Control				
Performance indicators						
Mean	242.57	242.89	-0.32	88	0.10	53
Median	238.82	238.95	-0.13	95	0.11	35
1st quartile	213.32	214.22	-0.90	64	0.10	45
3rd quartile	268.25	267.17	1.07	69	0.10	56
Standard deviation	40.46	39.96	0.50	56	0.12	28
Better performance						
Mean of best 25%	296.95	297.15	-0.20	95	0.13	24
Mean of best 50%	274.98	274.45	0.53	84	0.11	39
Mean of best 75%	259.00	258.96	0.05	98	0.12	31
Percentage of students with grade above 317 (goal of the Todos pela Educação NGO)	7.05	5.95	1.10	27	0.07	86
Percentage of students with grade above 300	10.93	10.15	0.78	52	0.08	81
Percentage of students with grade above 350	2.55	1.61	0.94	14	0.06	96
Percentage of students with grade above 275	20.94	21.13	-0.19	90	0.10	47
Worse performance						
Mean of the worst 25%	196.09	197.30	-1.20	40	0.13	22
Mean of the worst 50%	211.59	212.69	-1.10	52	0.14	13
Percentage of students with grade below 200	17.76	16.06	1.70	9	0.10	52
Percentage of students with grade below 175	2.75	2.69	0.06	86	0.05	99
Percentage of students with grade below 150	0.00	0.00	0.00	0	0.00	100
Average number of students who took the test	61	64	-4	50	0.09	60

DIFFERENCES IN PERFORMANCE IN MATHEMATICS BETWEEN THE TREATMENT AND CONTROL STRATA, 12th GRADE: PIAUÍ, BASELINE – 2014

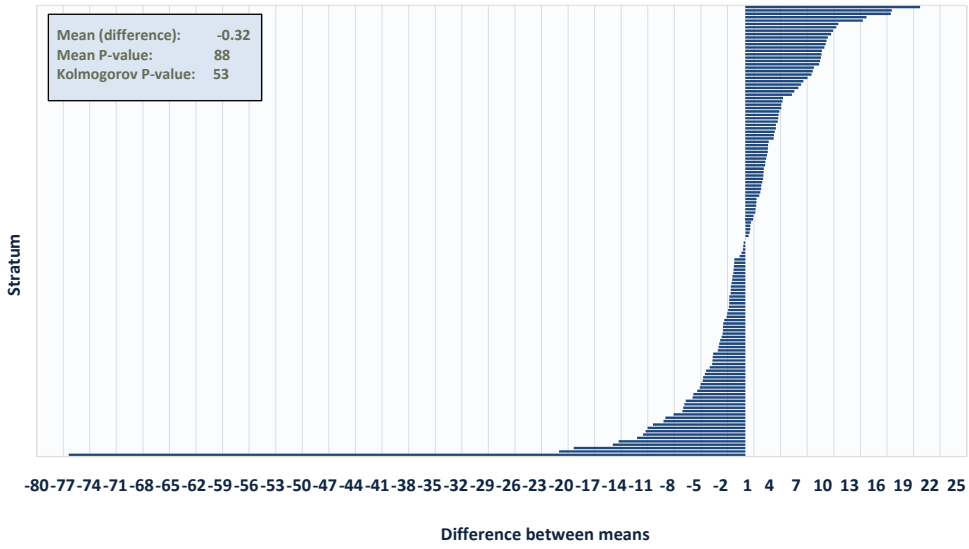
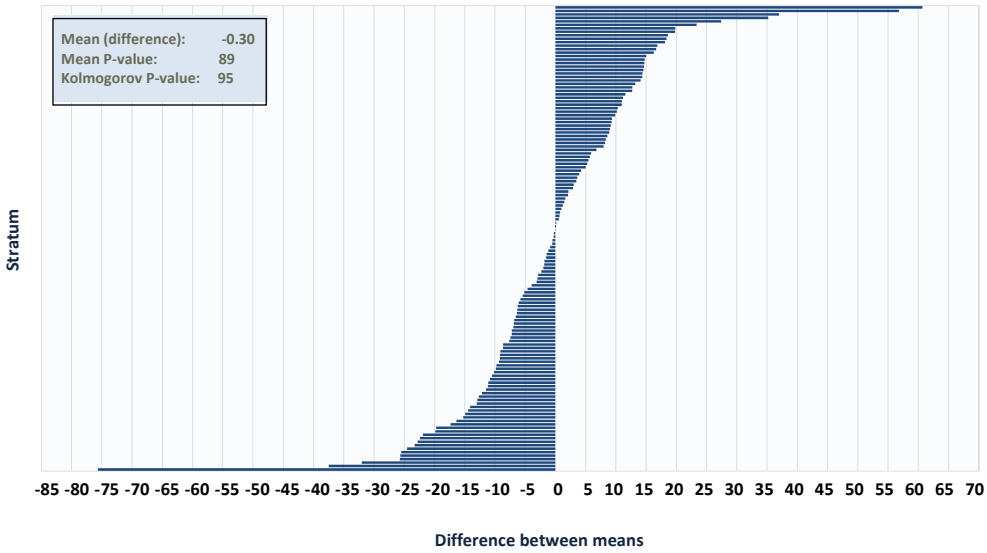


Table – Mean and Kolmogorov tests for the Portuguese Language baseline results in the Treatment and Control schools of the Program Jovem de Futuro: Piauí, 2014

Indicators	Mean (%)		Difference	P-value of T- test of equal means	Statistics of Kolmogorov test	P-value of Kolmogorov T-test
	Treatment	Control				
Performance indicators						
Mean	238.68	238.99	-0.30	89	0.06	95
Median	237.67	236.11	1.57	55	0.09	68
1st quartile	206.63	206.14	0.48	84	0.08	77
3rd quartile	268.20	269.86	-1.67	53	0.13	19
Standard deviation	43.45	44.49	-1.04	19	0.16	5
Better performance						
Mean among the best 25%	295.33	298.15	-2.83	26	0.14	14
Mean among the best 50%	273.96	275.27	-1.31	60	0.12	26
Mean among the best 75%	257.04	257.74	-0.70	77	0.09	67
Percentage of students with grade above 283 (goal of the Todos pela Educação NGO)	17.81	19.88	-2.06	18	0.13	20
Percentage of students with grade above 275	22.39	24.31	-1.92	26	0.11	39
Percentage of students with grade above 300	10.28	11.88	-1.61	17	0.13	18
Percentage of students with grade above 350	40.29	39.97	0.32	87	0.06	95

Worse performance						
Mean among the worst 25%	187.07	186.46	0.61	76	0.06	98
Mean among the worst 50%	205.43	204.58	0.85	69	0.06	99
Percentage of students with grade below 200	22.92	23.70	-0.78	63	0.06	94
Percentage of students with grade below 175	8.51	8.60	-0.09	92	0.10	48
Percentage of students with grade below 150	1.59	1.11	0.48	14	0.14	17
Average number of students who took the test	53	56	-3	53	0.09	69

DIFFERENCES IN PERFORMANCE IN PORTUGUESE LANGUAGE BETWEEN THE TREATMENT AND CONTROL STRATA, 12th GRADE: PIAUÍ, BASELINE – 2014



3. Randomization in the State of Pará

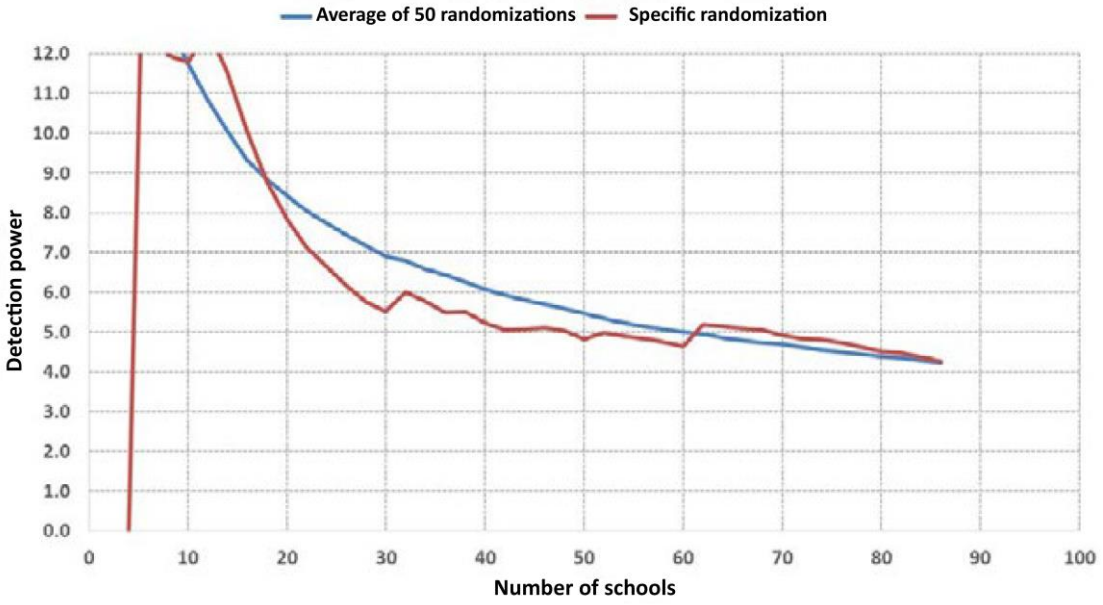
Tests

The method of pairing schools by predicting their performance evolution from 2014 to 2016 was considered the best one for schools in Pará. For the test, a group of 86 schools was considered and the following models were tested: (i) the prediction model, controlled by grades and the NSE of the schools, to pair up and select the treatment and control groups; and (ii) the random model selection the treatment and control groups.

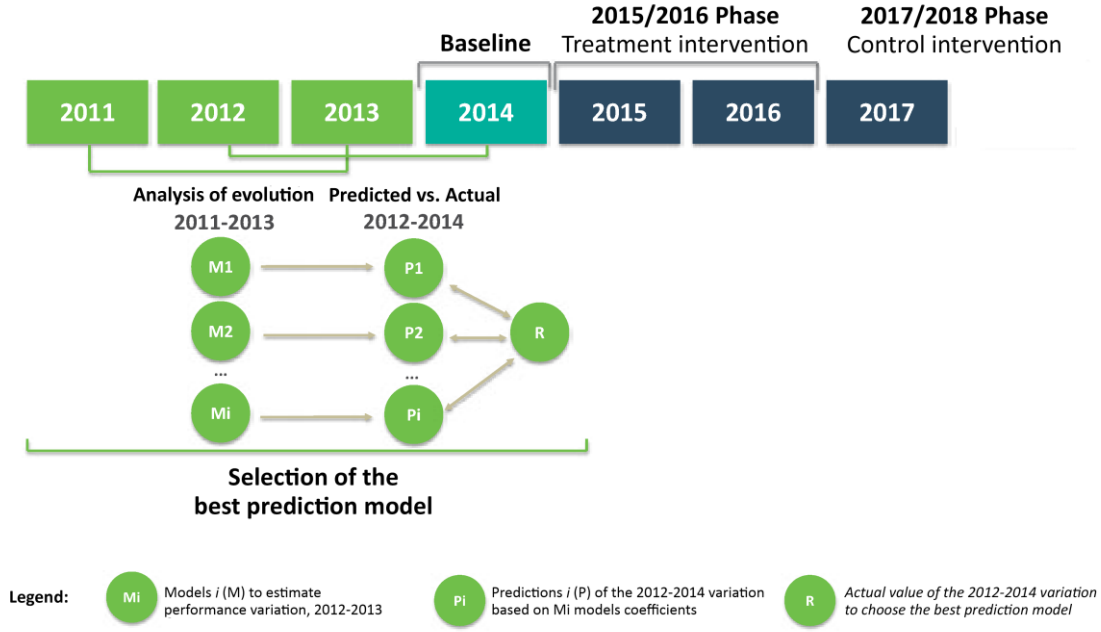
The criterion used to choose the best model was based on a lower impact detection power, that is, the chosen model would be the one most sensitive to capture the Program's impact. Thus, any future difference found between the control and treatment groups over and above the value of the detection power can be attributed to the Program's impact.

The prediction model (i) considered 42 control schools and 44 treatment schools, and its detection power was 4.2. The random model (ii) considered the same number of control and treatment schools and its detection power was 5.9. Thus, the model chosen to perform the randomization process was (i), the prediction model controlled by grades and the NSE of the schools.

DETECTION POWER BY NUMBER OF PAIRED SCHOOLS FROM THE CONTROL AND TREATMENT GROUPS: PARÁ – RANDOM PAIRING



THEORETICAL FRAMEWORK OF THE PREDICTION MODEL: THE TEST STRUCTURE, PARÁ



Delimitation of universes and strata

At this stage of redesigning the Program Jovem de Futuro (Phases 2015-2016 and 2017-2018), the agreement between Instituto Unibanco and the State

Department of Education resulted in a group of 87 schools eligible to participate in the Program.

Initially, these 87 schools were stratified into four groups, each comprising one municipality.

Due to the limited number of supervisors available in the state of Pará to provide support for vulnerable schools in the four municipalities chosen to participate in the Program, it was decided not to stratify schools by vulnerability in this state.

Thus, in Pará, the only level of stratification is by territory (level 1).

The final delimitation of the universe of schools for the randomization of control and treatment groups included 87 schools divided into 42 groups. The 87 schools were randomized using the prediction criterion.

Table – Universe of schools and Domain: Pará, Assessment by JF

Selection	Number of schools	Number of groups		
		Total	Pairs	Trios
Total number of eligible schools	87	42	39	3
Belém	47	23	22	1
Ananindeua	16	8	8	0
Marabá	13	6	5	1
Santarém	11	5	4	1

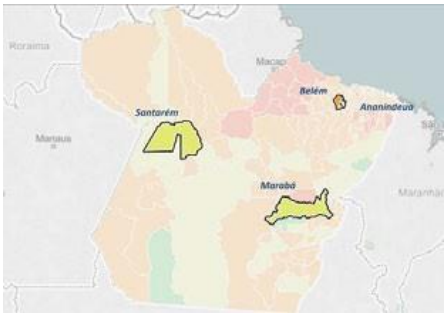


Table – Universe of schools and Domain: Pará, Assessment by JF

Selection	Number of schools			Number of groups		
	Total	Control	Treatment	Total	Pairs	Trios
Belém	47	23	24	23	22	1
Ananindeua	16	8	8	8	8	0
Marabá	13	6	7	6	5	1
Santarém	11	5	6	5	4	1
Total	**87**	**42**	**45**	**42**	**39**	**3**

The 2012-2014 model for estimating coefficients for the predictor of performance evolution between 2014 and 2016 generated the following statistics. Once the predicted value was calculated, schools were sorted within their strata and randomized.

Summary of the results

Regression statistics	
Multiple R	0.538
R squared	0.290
Adjusted R squared	0.263
Standard error	10.024
Observations	84

ANOVA

	GL	SS	MS	F-value	Significant F
Regression	3	3282.854	1.094.285	10.890	0.000
Residuals	80	8038.974	100.487		
Total	83	11321.827			

	Coefficients	Standard error	Stat T	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intersection	96.1622	21.3991	4.4938	0.0000	53.5766	138.7477	53.5766	138.7477
NSE	-0.5705	0.1825	-3.1256	0.0025	-0.9337	-0.2073	-0.9337	-0.2073
LP 2012	0.1425	0.1926	0.7400	0.4615	-0.2407	0.5257	-0.2407	0.5257
MT 2012	20.3065	9.2912	2.1856	0.0318	1.8163	38.7966	1.8163	38.7966

Validation

Tables with the values of the T-test for equal means and of the Kolmogorov-Smirnov test are shown below.

Table – Mean and Kolmogorov tests for the Mathematics baseline results in the Treatment and Control schools of the Program Jovem de Futuro: Pará, 2014

Indicators	Mean (%)		Difference	P-value of T test of equal means	Statistics of Kolmogorov test	P-value of Kolmogorov T test
	Treatment	Control				
Performance indicators						
Mean	237	239	-1,87	48	0,12	90
Median	233	237	-3,23	26	0,21	32
1st quartile	208	210	-1,59	51	0,12	93
3rd quartile	262	264	-2,41	49	0,16	64
Standard deviations	39	39	0,25	84	0,14	81
Better performance						
Mean among the best 25%	290	292	-1,59	69	0,17	54
Mean among the best 50%	269	271	-2,05	55	0,16	69
Mean among the best 75%	253	255	-2,13	49	0,14	82
Percentage of students with grade above 283 (goal of the Todos pela Educação NGO)	0,04	0,04	0,00	91	0,13	89
Percentage of students with grade above 300	0,08	0,08	0,00	90	0,11	95
Percentage of students with grade above 350	0,01	0,01	0,00	39	0,10	99
Percentage of students with grade above 275	0,17	0,18	-0,01	66	0,16	69
Worse performance						
Mean among worst 25%	193	195	-1,24	46	0,24	19
Mean among worst 50%	207	209	-1,93	35	0,15	69
Percentage of students with grade below 200	0,19	0,19	0,00	79	0,16	67
Percentage of students with grade below 175	0,03	0,02	0,01	25	0,20	39
Percentage of students with grade below 150	0,00	0,00	0,00	0	0,00	100
Average number of students who took the test	71	100	-29	3	0,23	20

DIFFERENCES IN PERFORMANCE IN MATHEMATICS BETWEEN THE TREATMENT AND CONTROL STRATA, 12th GRADE: PARÁ, BASELINE – 2014

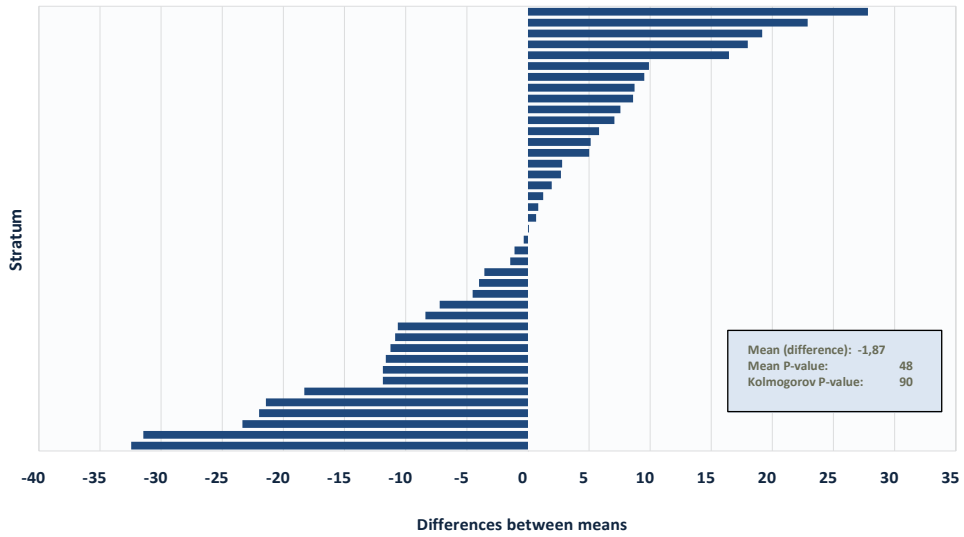
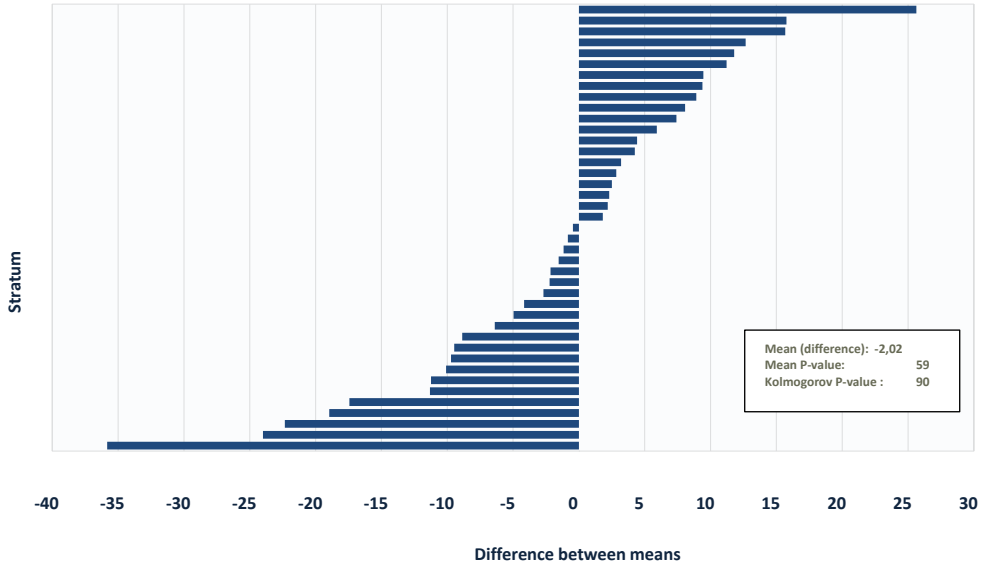


Table – Mean and Kolmogorov tests for the Portuguese Language baseline results in the Treatment and Control schools of the Program Jovem de Futuro: Pará, 2014

Indicators	Mean (%)		Difference	P-value of T test of equal means	Statistics of Kolmogorov test	P-value of Kolmogorov T test
	Treatment	Control				
Performance indicators						
Mean	226	228	-2,02	59	0,12	90
Median	222	226	-4,02	38	0,13	87
1st quartile	190	191	-1,40	70	0,15	75
3rd quartile	259	262	-2,48	60	0,14	77
Standard deviation	48	48	-0,18	89	0,14	79
Better performance						
Mean among the best 25%	290	292	-1,48	76	0,10	99
Mean among the best 50%	265	267	-2,28	62	0,14	78
Mean among the best 75%	245	248	-2,79	51	0,15	75
Percentage of students with grade above 283 (goal of the Todos pela Educação NGO)	0,16	0,16	0,00	98	0,12	91
Percentage of students with grade above 275	0,19	0,19	0,00	94	0,12	92
Percentage of students with grade above 300	0,09	0,09	0,00	80	0,15	74
Percentage of students with grade above 350	0,31	0,34	-0,03	35	0,14	78
Worse performance						
Mean among the worst 25%	170	170	-0,30	91	0,15	72
Mean among the worst 50%	188	191	-2,09	51	0,14	83

Percentage of students with grade below 200	0,35	0,33	0,01	60	0,11	97
Percentage of students with grade below 175	0,18	0,16	0,02	40	0,19	43
Percentage of students with grade below 150	0,04	0,04	0,00	80	0,09	99
Average number of students who took the test	71	100	-29	3	0,23	20

DIFFERENCES IN PERFORMANCE IN PORTUGUESE LANGUAGE BETWEEN THE TREATMENT AND CONTROL STRATA, 12th GRADE: PARÁ, BASELINE – 2014



4. Randomization in the State of Espírito Santo

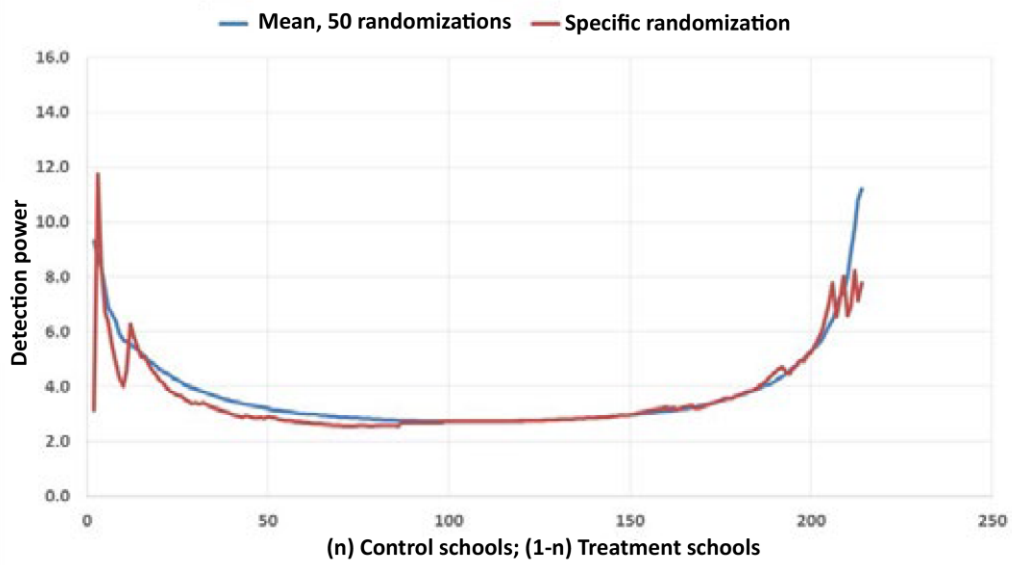
Tests

The method of pairing schools by predicting their performance evolution from 2014 to 2016 was considered the best one for Espírito Santo. For the test, a group of 216 schools was considered and the following models were tested: (i) the prediction model, controlled by grades and the NSE of the schools, to pair up and select the treatment and control groups; and (ii) the random model to select the treatment and control groups.

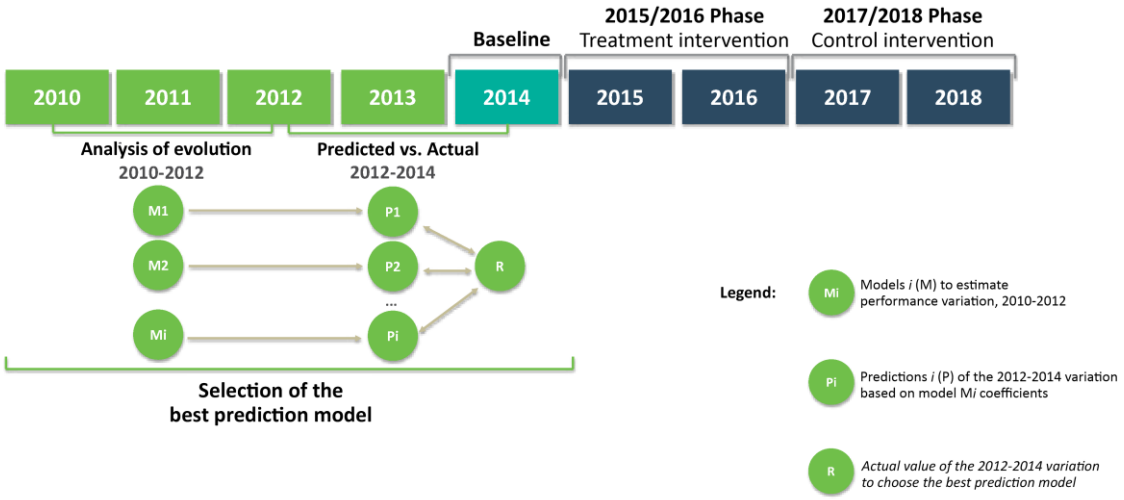
The criterion used to choose the best model was based on a lower impact detection power, that is, the chosen model would be the one most sensitive to capture the Program's impact. Thus, any future difference found between the control and treatment groups over and above the value of the detection power can be attributed to the Program's impact.

The prediction model (i) considered 60 control schools and its detection power was 2.7. The random model (ii) considered the same number of control schools and its detection power was 3.1. Thus, the model chosen to perform the randomization process was (i), the prediction model controlled by grades and the NSE of the schools.

DETECTION POWER IN THE 2012-2014 VARIATION OF THE MEAN GRADE (PORTUGUESE LANGUAGE & MATHEMATICS) ACCORDING TO THE NUMBER OF CONTROL SCHOOLS: ESPÍRITO SANTO, 12TH GRADE



THEORETICAL FRAMEWORK OF THE PREDICTION MODEL: THE TEST STRUCTURE, ESPÍRITO SANTO



Because of the availability of data, the test models in Espírito Santo were developed for the performance variation between 2010 and 2012, controlled by the NSE and by the grades of 2011 and 2010.

Thus, the final prediction model (2014-2016) will be controlled by the NSE and the grades of 2012, 2013 and 2014.

Delimitation of universes and strata

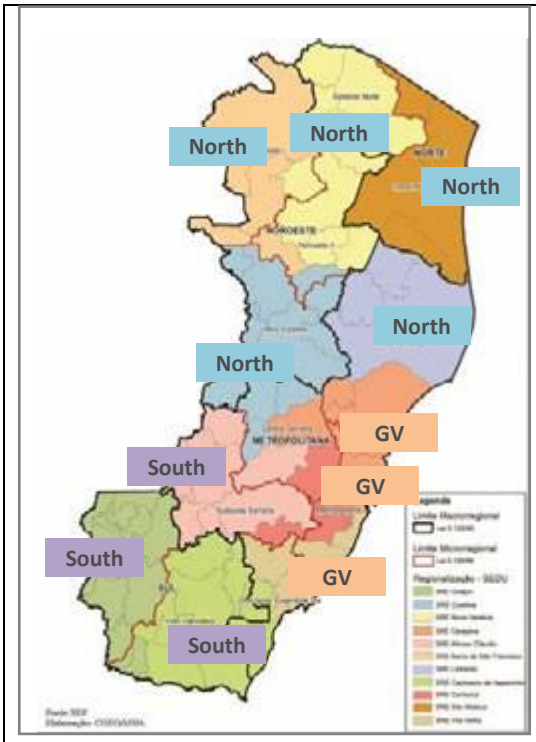
In this deployment stage of the Program Jovem de Futuro (Phases 2015-2016 and 2017-2018), the agreement between Instituto Unibanco and the State Department of Education resulted in a group of 222 schools eligible to participate in the Program.

The criterion used by the Department to delimit the universe was that a school had to have at least 120 students enrolled in regular high school in 2015 (the year the Program's first phase began).

Initially, these 222 schools were stratified into four groups, each comprising a number of Regional Offices of education.

Table – Universe of Schools and Domain: Espírito Santo, Assessment by JF

Selection	Number of schools	Number of groups		
		Total	Groups of 3	Groups of 4
Total no. of schools on file	288	–	–	–
With at least 120 high school students in 2015	222	70	59	11
North	69	22	19	3
Greater Vitória	92	29	24	5
South	60	19	16	3
No grades in 2014	1	–	–	–



Regional Office	Region
SRE BARRA DE SÃO FRANCISCO	North
SRE COLATINA	North
SRE LINHARES	North
SRE NOVA VENÉCIA	North
SRE SÃO MATEUS	North
SRE CARAPINA	Greater Vitória
SRE CARIACICA	Greater Vitória
SRE VILA VELHA	Greater Vitória
SRE AFONSO CLÁUDIO	South
SRE CACHOEIRO DE ITAPEMIRIM	South
SRE COMENDADORA JUREMA MORETZ SOHN	South

Of the 222 schools selected to compose the universe, one had no grades. As a result, the baseline could not be determined and, therefore, that school was not included in the assessment, reducing the group of assessed schools to 221. The school removed from the assessment is listed below, and it is still being decided whether this gradeless school will enter the first or the second phase of the Program's intervention:

- 32037163 – EEEFM Francisco Nascimento (Greater Vitória);

The number of control schools was also determined by the Department of Education in the agreement with Instituto Unibanco. In keeping with the stratification criteria, the Program will be implemented in the entire system and there will be 70 control schools so as to make it possible to assess its impact.

Table – Level 1 stratification: Espírito Santo, Assessment by JF

Selection	Number of schools	Number of groups			Domain	
		Total	Groups of 3	Groups of 4	T	C
Total no. of schools on file	288	–	–	–	–	–
With at least 12 high school students in 2015	222	70	59	11	151	70
North	69	22	19	3	47	22
Greater Vitória	92	29	24	5	63	29
South	60	19	16	3	41	19
No grade in 2014	1	–	–	–	–	–

For the second level of stratification (specifying the vulnerable schools), the Department decided that the domain would be limited to 30 schools.

- North – 5 trios, 15 schools;
- Greater Vitória – 2 trios, 6 schools; and
- South – 3 trios, 9 schools.

Table – Level 2 stratification – Vulnerability criteria: Espírito Santo

School code	School	Region	Regional Office (GRE)
32015631	EEEM NOSSA SENHORA DE LOURDES	North	SRE NOVA VENÉCIA
32075936	EEEFM PROFª ANTONIETA BANHOS FERNANDES	North	SRE LINHARES
32013728	EEEFM SEBASTIANA GRILO	North	SRE COLATINA
32007876	EEEFM SÃO GABRIEL DA PALHA	North	SRE NOVA VENÉCIA
32019459	EEEFM NESTOR GOMES	North	SRE SÃO MATEUS
32010699	EEEFM PROFª NÉA MONTEIRO COSTA	North	SRE COLATINA
32000499	EEEFM DERMEVAL LEITE RIBEIRO	North	SRE BARRA DE SÃO FRANCISCO
32020333	EEEFM ERMENTINA LEAL	North	SRE LINHARES
32001916	EEEFM PROF ASCENDINA FEITOSA	North	SRE BARRA DE SÃO FRANCISCO
32078528	EEEM SANTINA MOROSINI CUPERTINO	North	SRE LINHARES
32007175	EEEFM ALARICO JOSÉ DE LIMA	North	SRE NOVA VENÉCIA
32009402	EEEFM DR. JONES DOS SANTOS NEVES	North	SRE COLATINA
32005652	EEEFM ANTONIO DOS SANTOS NEVES	North	SRE NOVA VENÉCIA
32013906	EEEFM JANUARIO RIBEIRO	North	SRE COLATINA
32030584	EEEFM EURICO SALLES	North	SRE COLATINA
32063199	EEEFM ELZA LEMOS ANDREATTA	GV	SRE CARAPINA

32065043	EEEFM EWERTON MONTENEGRO GUIMARÃES	GV	SRE CARIACICA
32076410	EEEFM DR. JOSÉ MOYSES	GV	SRE CARIACICA
32079230	EEEM MARIO GURGEL	GV	SRE VILA VELHA
32043686	EEEM DR SILVA MELLO	GV	SRE VILA VELHA
32034016	EEEFM ANA LOPES BALESTRERO	GV	SRE CARIACICA
32052405	EEEFM LIONS SEBASTIÃO PAIVA VIDAURRE	South	SRE CACHOEIRO DE ITAPEMIRIM
32045360	EEEFM PROFESSOR PEDRO SIMÃO	South	SRE COMENDADORA JUREMA MORETZ SOHN
32049242	EEEFM P AFONSO BRAZ	South	SRE COMENDADORA JUREMA MORETZ SOHN
32025920	EEEFM AFONSO CLÁUDIO	South	SRE AFONSO CLÁUDIO
32048459	EEEFM BERNARDO HORTA	South	SRE COMENDADORA JUREMA MORETZ SOHN
32046103	EEEFM SIRENA REZENDE FONSECA	South	SRE COMENDADORA JUREMA MORETZ SOHN
32046022	EEEFM JOSÉ CORRENTE	South	SRE COMENDADORA JUREMA MORETZ SOHN
32053622	EEEFM PROFESSOR DOMINGOS UBALDO	South	SRE CACHOEIRO DE ITAPEMIRIM
32046197	EEEFM PROF CELIA TEIXEIRA DO CARMO	South	SRE COMENDADORA JUREMA MORETZ SOHN

The final delimitation of the universe of schools for the randomization of the control and treatment groups included 221 schools divided into 70 groups:

Table – Level 1 stratification: Espírito Santo, Assessment by JF

Region	Schools			Groups		
	Total	Control	Treatment	Non vulnerable trios	Non vulnerable quartets	Vulnerable trios
North	69	22	47	14	3	5
Greater Vitória	92	29	63	22	5	2
South	60	19	41	13	3	3
No grade in 2014	1	–	–	–	–	–
Total	**222**	**70**	**151**	**49**	**11**	**10**

Of the 221 schools taking part in the randomization process for selection of the control and treatment groups, 188 were paired according to the prediction of performance variation for the 2014-2016 period; 20 were paired according to their vulnerability; 3 were randomly paired. The school without grades in 2014 remains awaiting definition.

The 2012-2014 model for estimating the coefficients for the predictor of performance evolution between 2014 and 2016 generated the following

statistics. Once the predicted value was calculated, schools were sorted within their strata and randomized.

SUMMARY OF RESULTS

Regression statistics	
Multiple R	0.547
R squared	0.299
Adjusted R squared	0.275
Standard error	10.350
Observations	215

ANOVA

	GL	SS	MS	F-value	Significant F
Regression	7	94446.940	1349.563	12.599	0.000
Residuals	207	22172.740	107.115		
Total	214	31619.680			

	Coefficients	Standard error	Stat T	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intersection	61.473	15.708	3.913	0.000	30.505	92.441	30.505	92.441
NSE	-0.257	1.359	-0.189	0.850	-2.936	2.422	-2.936	2.422
LP 2010	-0.107	0.116	-0.923	0.357	-0.336	0.122	-0.336	0.122
LP 2011	0.037	0.123	0.297	0.767	-0.207	0.280	-0.207	0.280
LP 2012	-0.286	0.110	-2.601	0.010	-0.503	-0.069	-0.503	-0.069
MT 2010	0.297	0.104	2.852	0.005	0.092	0.503	0.092	0.503
MT 2011	0.212	0.119	1.778	0.077	-0.023	0.447	-0.023	0.447
MT 2012	-0.385	0.099	-3.879	0.000	-0.581	-0.189	-0.581	-0.189

Validation

Tables with the values of the T-test for equal means and of the Kolmogorov-Smirnov test are shown below.

Table – Mean and Kolmogorov tests for the Mathematics baseline results in the Treatment and Control schools of the Program Jovem de Futuro: Espírito Santo, 2014

Indicators	Mean (%)		Difference	P-value of T test of equal means	Statistics of Kolmogorov test	P-value of Kolmogorov T test
	Treatment	Control				
Performance indicators						
Mean	275	278	-2,55	35	0,10	74
Median	275	278	-2,38	42	0,10	69
1st quartile	242	245	-2,84	36	0,12	52
3rd quartile	308	311	-3,37	25	0,15	21
Standard deviation	47	48	-0,36	62	0,10	73
Better performance						
Mean of best 25%	335	338	-2,69	32	0,11	66
Mean of best 50%	313	316	-2,76	31	0,13	37
Mean of best 75%	296	298	-2,67	33	0,13	44
Percentage of students with grade above 317 (goal of the Todos pela Educação NGO)	0,21	0,24	-0,03	9	0,14	28
Percentage of students with grade above 300	0,32	0,34	-0,02	31	0,12	52
Percentage of students with grade above 350	0,07	0,08	-0,01	45	0,09	83
Percentage of students with grade above 275	0,50	0,53	-0,02	35	0,10	68
Worse performance						
Mean of worst 25%	216	218	-1,96	48	0,10	73
Mean of worst 50%	238	240	-2,32	41	0,11	59
Percentage of students with grade below 200	0,08	0,08	0,00	70	0,07	98
Percentage of students with grade below 175	0,02	0,02	0,00	82	0,07	97
Percentage of students with grade below 150	0,00	0,00	0,00	62	0,02	100
Average number of students who took the test	96	85	11	21	0,18	10

DIFFERENCES IN PERFORMANCE IN MATHEMATICS BETWEEN THE TREATMENT AND CONTROL STRATA, 12th GRADE: ESPÍRITO SANTO, BASELINE – 2014

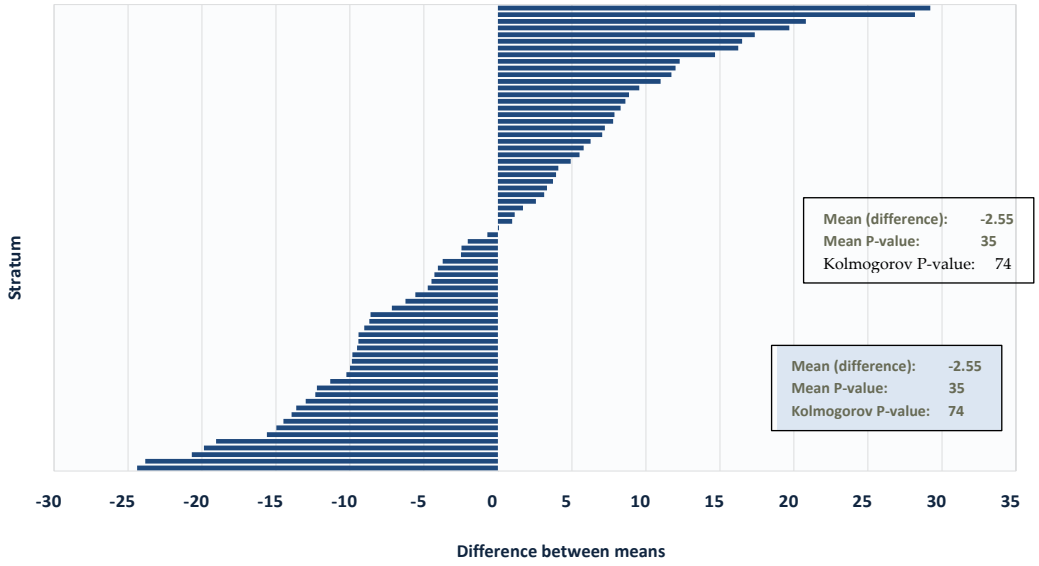
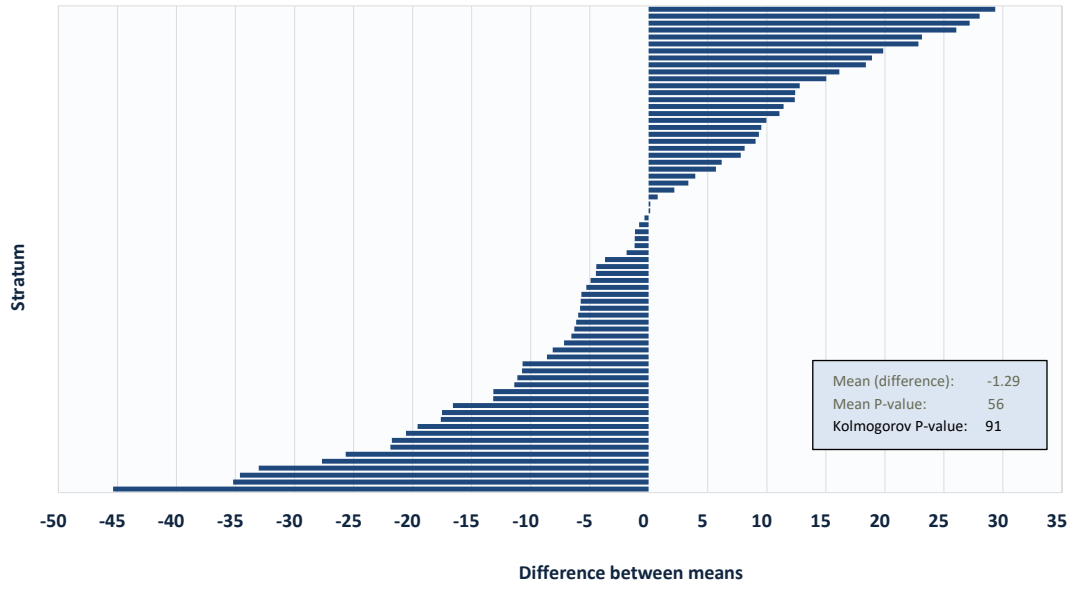


Table – Mean and Kolmogorov tests for the Portuguese Language baseline results in the Treatment and Control schools of the Program Jovem de Futuro: Espírito Santo, 2014

Indicators	Mean (%)		Difference	P-value of T-test of equal means	Statistics of Kolmogorov test	P-value of Kolmogorov T-test
	Treatment	Control				
Performance indicators						
Mean	266	267	-1,29	56	0,08	91
Median	268	270	-1,24	61	0,09	82
1st quartile	235	236	-1,65	55	0,09	85
3rd quartile	298	301	-2,69	22	0,15	24
Standard deviation	46	46	-0,49	51	0,14	35
Better performance						
Mean of best 25%	322	324	-1,72	38	0,10	68
Mean of best 50%	303	304	-1,72	40	0,12	53
Mean of best 75%	286	288	-1,56	47	0,09	83
Percentage of students with grade above 283 (goal of the Todos pela Educação NGO)	0,39	0,41	-0,02	29	0,12	51
Percentage of students with grade above 275	0,45	0,47	-0,02	28	0,13	37
Percentage of students with grade above 300	0,25	0,27	-0,02	19	0,17	14
Percentage of students with grade above 250	0,65	0,65	0,00	93	0,09	85
Worse performance						
Mean of worst 25%	208	208	-0,26	92	0,07	99
Mean of worst 50%	230	231	-0,86	72	0,08	95
Percentage of students with grade below 200	0,10	0,10	0,00	95	0,06	100
Percentage of students with grade below 175	0,04	0,04	0,00	97	0,07	97
Percentage of students with grade below 150	0,01	0,01	0,00	82	0,07	98
Average number of students who took the test	96	85	11	21	0,18	10

DIFFERENCES IN PERFORMANCE IN PORTUGUESE LANGUAGE BETWEEN THE TREATMENT AND CONTROL STRATA, 12th GRADE: ESPÍRITO SANTO, BASELINE – 2014



Reference

BAKER, Judy (2000) Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners. The World Bank, Washington. Available at: <https://openknowledge.worldbank.org/bitstream/handle/10986/13949/207450PUB0Repl0xtop0146970same0info.pdf?sequence=1>.