

DIL Research Assistant Report

To: Policymakers of Chicago

From: Pedro Huet Date: 09/21/2023

RE: School subsidy randomized control trial analysis and impact evaluation report.

Executive Summary

The literature has argued that subsidies and transfer programs could be effective to increase education and development in communities, so it makes sense to perform experiments to verify the accuracy and magnitude of said variable. An RTC, which employed pairwise randomization treatment assignment, of a proposed subsidy program was conducted in 165 schools across a sample of almost 10,000 elementary-level students located in 2 districts from 2010 to 2012. Two follow-up surveys were conducted after the start of the intervention, one the year after the end of the program, and the second 2 years later. Performing a series of logistic regression models, the data supports the hypothesis that school subsidies reduce the likelihood of school evasion, child pregnancies and marriages during the length of the program, as well as some years afterwards; although the magnitude of the effects are comparatively small when considering other variables. Further, the results of a simple panel difference-in-difference model favors the argument that the subsidy has both short-term and long-term effects on reducing school dropouts. However, the effect of school subsidies wanes when considering that child pregnancies and marriages could also be causes of school dropouts. The conclusion is that school subsidies seem to have a small effect in reducing school evasion, yet the feasibility of this approach isn't as convincing when considering there could be alternative causes to school dropouts which could be employed at a lower cost.

Background & Research

The bolstering of education as a fundamental component of social development has been a topic of great relevance in the last 3 decades. It has also been supported by the literature and multiple studies (Speil et al., 2018; Öztürk et al., 2022). Given this situation, it follows that there could be some policy alternatives to encourage or aid youths' access to this resource. Among the different paths that have been argued to bolster education among communities, many authors have argued in favor of utilizing subsidies, conditional cash/in-kind transfers (Ullmann et al., 2021;). The argument is that giving children and their families access to additional resources for pursuing their studies adds an additional incentives for the recipients to continue pursuing their education.

This argument is enticing, yet the policy it supports seems to implicate the use of a large amount of resources (Loser et al., 2021), meaning it could have an important toll on public finances. In order to assess if subsidies and cash or in-kind transfers are worth the investment, it would make



sense to study the magnitude and reliability of their impact. As a contribution to this task, this report proposes a statistical analysis based on an RTC designed experiment based on a subsidies program targeted towards the education of students.

Study Methodology

The study consist of evaluating an education program that subsidizes the cost of education for students through in-kind transfers. The study consisted of a program that lasted for three years, from 2010 to 2012, and it targeted children enrolled in public schools. Throughout this time period, the subsidies were delivered at the beginning of every school year to the cohort of students enrolled in grade 6 at the start of the program.

To evaluate its effectiveness, an experiment was designed using pairwise randomization to assign schools within two districts to a treatment or control group. The main unit of analysis are the children whose schools did or did not receive the subsidies. All children enrolled in the treatment schools received the subsidy for the three years, conditional that they were still enrolled in school. To conduct the study, baseline data was collected for the schools and cohorts, and two follow-up surveys were conducted: the first one, on the year after the last subsidy was offered to students (year 3), and another two years later (year 5).

With regards to performance metrics for the program, the study considered school evasion, if a student dropped out of his or her educational formation, as the primary outcome of interest. It considered teen pregnancy and teen marriage as secondary variables of interest.

Based on the available data, it would seem reasonable to choose a series of variables that likely have a larger effect on affecting school evasion. The first and most relevant variable for the study is treatment (dichotomous variable), if the student was in a school that gave him or her access to the subsidy, as this result allows us to approach the effectiveness of subsidies and transfer programs. A second variable would be the sex of the student, as its possible that male or female students, due to social and contextual factors, may have a larger tendency to dropout of their education. A third variable is the year of birth of students, since younger students tend to have less motivation of dropping out of school (they are usually less qualified at finding alternative venues of employment than older students). Since this variable is categorized by the year of birth of the students receiving the subsidies (between 1985 and 1992), the variable can be treated as ordinal (continuous, for our purposes). Finally, we also consider the month in which the visit was made to students (nominal variable), since months closer to the start and end of the school year (June, July, August, and December) are likely to be periods in which the students are more prone to end their studying experiencing (students could be more likely to want to finish their semesters due to a process of inertia).

With regards to the dependent variables, school dropouts (if the student left school), teen pregnancies (if a student or his partner experienced a pregnancy) and marriages (if the student got



married) are measured as dichotomous variables (from 0 to 1). These variables will be measured both during the year 3 since the start of the intervention and on year 5 of it occurring.

The study consists of a series of a maximum likelihood logarithmic models estimating the short-term and long-term effects of the intervention, as well as a difference-in-difference panel linear model. The advantage of using a non-linear model is that it is measuring the probabilities of an outcome occurring, in this case, a student dropping out, rather than a continuous quantity (i.e. "the degree in which a student drops out"). Because of this, a non-linear model like a logistic model is more versatile than a regular linear probability model, meaning it'll likely make a more useful estimation. The main model can be summarized by the following equation.

School Dropouts =
$$B_0 + B_1$$
(treatment) + B_2 (sex) + B_3 (year of birth) + B_4 (month of visit) + u

With regards to the linear fixed effects difference-in-difference model, since the panel data only consists of two years, it doesn't make sense to use more sophisticated models (i.e. Difference-in-Difference-in-Differences model or Multiple-Timming-Fixed Effects Models), as there isn't enough data to obtain noticeably different estimations. Perhaps the main problem on the outset of this study is that we cannot verify the Parallel Trends Assumption: it isn't clear that the individuals who were treated on the third year of the program followed the same trend as those who were untreated during this period. Although it would have been preferable to perform this analysis just before the start of the intervention to have more certainty of both groups being comparable, an analysis of this sort can still be insightful to observe if there were differences between the effects of the subsidies at the 3rd year of the intervention and the 5th year. The panel model can be summarized by the following equation.

School Dropouts =
$$B_0 + B_1$$
(treatment) + B_2 (long-term) + B_3 (treatment * long-term) + u

Key Findings

Before starting the analysis, it is important to make sure that the Randomized Control Trial worked in the correct manner. Table 1 shows that a multivariate regression model that includes the variables that could have been affecting the school's assignment does not find much evidence of bias in the assignment. Aside from the number of latrines the schools have, none of the other variables related to the schools (number of teachers and students, average student scores, rural location, etc.) seem to be associated to a significance level of above 10% with being assigned to the treatment group; while none of the variables reaches a confidence level of above 5%. Because of this, it seems that, in general, the baseline characteristics seem balanced.



Table 1

treatment	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
district	2465828	.4286512	-0.58	0.565	-1.086724	.5935582
location	8957887	.6688227	-1.34	0.180	-2.206657	.4150797
n_teachers	0612822	.1076023	-0.57	0.569	2721788	.1496143
n_teachers_fem	.0043786	.0850735	0.05	0.959	1623624	.1711195
female_head_teacher	.298549	.611635	0.49	0.625	9002335	1.497332
n_students_fem	0056233	.005323	-1.06	0.291	0160562	.0048097
n_students_male	.0016141	.0054949	0.29	0.769	0091557	.0123839
n_schools_2km	040937	.1052616	-0.39	0.697	2472459	.165372
av_teacher_age	.0807291	.0601273	1.34	0.179	0371183	.1985764
av_student_score	.0057991	.0070456	0.82	0.410	0080101	.0196083
n_latrines	.074235	.038603	1.92	0.054	0014256	.1498955
_cons	-2.694787	3.411237	-0.79	0.430	-9.38069	3.991115

With regards to the short-term effects of the subsidies, Tables 2, 3 and 4 show the possible short-term effects of receiving subsidies on children's education. Table 2 shows that treatment into the group after 3 years has a negative and statistically significant effect, at a significance level of above 1%, in the student dropping out. When seeing the magnitude, its size is around 2.2%, meaning that, if all other variables remain equal, a student that is in a school that receives subsidies is 2.2% less likely to drop-out of school. A student's year of birth also show statistically significant negative effects on school evasion, meaning that, for each year a student is younger, all else equal, he or she is almost 6% less likely to dropout. When observing sex, we observe that being a female student, all else equal, seems to be associated with a 8.6% probability of being more prone to dropout of school. Finally, the students who were surveyed in October were less prone to dropout by around 9%, all else equal. The results of these variables are all significant at a level of above 5%. Figure 1 offers a visual summary of the magnitude and significance level of the models' findings.



٦.	•		
a	h	le	,
а	.,	ı	

Logistic regression Number of obs = 9,004LR chi2(7) = 787.01Prob > chi2 = 0.0000Log likelihood = -3222.5966Pseudo R2 = 0.1088______ dropout | Coefficient Std. err. z P>|z| [95% conf. interval] ______ -.3403776 -.0856488
 yob | -.5882476
 .0231562
 -25.40
 0.000
 -.633633
 -.5428623

 sex | .8022253
 .0684522
 11.72
 0.000
 .6680614
 .9363893
 visit month | 9 | -.4789905 .2690345 -1.78 0.075 -1.006288 .0483074 10 | -.8116111 .2812518 -2.89 0.004 -1.362854 -.2603678 11 | -.1588488 .2698362 -0.59 0.556 -.6877179 .3700204 12 | -.1925838 .3149806 -0.61 0.541 -.8099344 .4247669 _cons | 1168.343 46.05379 25.37 0.000 1078.079 1258.607

Average marginal effects

Number of obs = 9,004

Model VCE: OIM

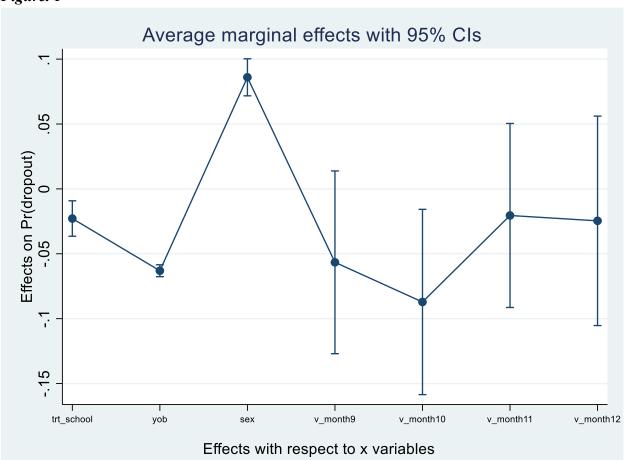
Expression: Pr(dropout), predict()

dy/dx wrt: trt_school yob sex 9.visit_month 10.visit_month 11.visit_month 12.visit_month

	 	Delta-metho	 od			
	dy/dx	std. err.	Z	P> z	[95% conf	. interval]
trt_school	022835	.0069611	-3.28	0.001	0364786	0091915
yob	0630602	.0023503	-26.83	0.000	0676666	0584538
sex	.0859986	.0072841	11.81	0.000	.071722	.1002753
	I					
visit_month	I					
9	0565799	.0359211	-1.58	0.115	1269839	.0138242
10	0871516	.0364265	-2.39	0.017	1585462	015757
11	0204904	.0361792	-0.57	0.571	0914004	.0504196
12	0246181	.0411715	-0.60	0.550	1053128	.0560767







Tables 3 and 4 offer similar results. They show that, as with school evasion, it would also seem that the subsidies program has negative and statistically significant effects (of a significance level of above 95%) towards experiencing teen pregnancy and marriages, yet it also seems relatively small (around 2% reduction).

Table 3						
Logistic regre	ssion				Number of ob	s = 4,328
					LR chi2(6)	
T 1 / 1 - 1 / 1 1	1220 7050				Prob > chi2	
Log likelihood	1 = -1328.7058	3			Pseudo R2	= 0.1489
pregnt	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
+						
					4573321	
yop	7762106	.040134	-19.34	0.000	8548718	6975493
sex	0	(omitted)				
1						
visit_month						
9	.2279757	.530794	0.43	0.668	8123613	1.268313
10	3226972	.5488073	-0.59	0.557	-1.39834	.7529454
11	.5934452	.5323371	1.11	0.265	4499164	1.636807
12	.1519605	.614408	0.25	0.805	-1.052257	1.356178

Average marginal effects Number of obs = 4,328

Model VCE: OIM

Expression: Pr(pregnt), predict()

dy/dx wrt: trt_school yob sex 9.visit_month 10.visit_month 11.visit_month 12.visit_month

Delta maked

			Delta-method	l			
	1	dy/dx	std. err.	Z	P> z	[95% conf.	interval]
trt_school		0228837	.0091636	-2.50	0.013	040844	0049233
yob		0693375	.0034071	-20.35	0.000	0760153	0626597
sex		0	(omitted)				
visit_month							
9		.0184215	.0398947	0.46	0.644	0597706	.0966136
10		0216869	.0404933	-0.54	0.592	1010523	.0576784
11		.0539865	.0403937	1.34	0.181	0251837	.1331566
12		.0119742	.0473682	0.25	0.800	0808657	.104814



			le	4
	•	n	10	4
_	a	w	ı	т

Logistic regre	Number of ob	•				
Log likelihood	= -1414.583				Prob > chi2 Pseudo R2	= 0.0000
married			z		[95% conf.	interval]
·	3384766					133638
yob	7737201	.0377909	-20.47	0.000	8477889	6996513
sex	2.625267	.1480263	17.74	0.000	2.335141	2.915394
visit_month						
9	.0700375	.5211146	0.13	0.893	9513283	1.091403
10	700827	.5449954	-1.29	0.198	-1.768998	.3673443
11	.3370815	.5230385	0.64	0.519	6880551	1.362218
12	.2403885	.5833123	0.41	0.680	9028827	1.38366
_cons	1534.292 	75.1138	20.43	0.000	1387.072	1681.513

. margins, dydx(*)

Average marginal effects

Model VCE: OIM

Expression: Pr(married), predict()

dy/dx wrt: trt_school yob sex 9.visit_month 10.visit_month 11.visit_month 12.visit_month

Number of obs = 8,791

With regards to long-term effects, Tables 5, 6 and 7 show that the results remained similar: if a student recieves subsidies, his or her dropout rate decreased by around 5%, which is slightly higher than what was found in the short term? at a level of statistical significante of above 5%. Perhaps the most important difference in the long-term can be seen in Table 7: the negative effect of subsidies on teen marriage seems to decrease noticeably after 5 years of the program's start, which is reflected in the fact that the significance level of this variable decrease to below 5%, but comfortably above 10%, and the magnitude of the effect decreased to around 1%.

Table 5

Logistic regression	Number of obs = $8,990$
	LR chi2(5) = 858.02
	Prob > chi2 = 0.0000
Log likelihood = -4538.2994	Pseudo R2 = 0.0864

dropout	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
trt_school yob sex	3251957 4891332 .7239195	.0526112 .0190829 .0546015	-6.18 -25.63 13.26	0.000	4283118 5265351 .6169025	2220795 4517313 .8309366
visit_month 5	.2983619	.0757346	3.94 2.36	0.000	.1499247	.4467991 .2965689
_cons	971.4962	37.95153	25.60	0.000	897.1126	1045.88

Average marginal effects Number of obs = 8,990

Model VCE: OIM

Expression: Pr(dropout), predict()

dy/dx wrt: trt_school yob sex 5.visit_month 7.visit_month

I	1	Delta-method	l			
 +	dy/dx	std. err.	Z	P> z	[95% conf.	interval]
trt_school	0537056	.0086365	-6.22	0.000	0706327	0367784
yob	0807796	.002787	-28.98	0.000	0862421	0753171
sex	.1195542	.008775	13.62	0.000	.1023554	.1367529
visit_month						
5	.0490473	.0123383	3.98	0.000	.0248647	.0732299
7	.0258428	.0107618	2.40	0.016	.0047501	.0469355

Table 6

Logistic regression Number of obs = $3,97$ LR chi2(4) = 555.2								
Log likelihoo	od = -2058.9237	Prob > chi2 Pseudo R2	= 0.0000 = 0.1188					
pregnt	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]		
_	2683974 687464 0	.0327906						
visit_month 5 7 _cons	.0621545 .3013065	.1019705	0.54 2.95 20.95	0.003	1631722 .1014479	.2874813 .5011651 1494.617		

Average marginal effects Number of obs = 3,974

Model VCE: OIM

Expression: Pr(pregnt), predict()

dy/dx wrt: trt_school yob sex 5.visit_month 7.visit_month

Note: dy/dx for factor levels is the discrete change from the base level.

Table 7



Logistic regression

L	og likelihood	d = -2536.9362	2		LR chi2(5) Prob > chi2 Pseudo R2	= 0.0000
_		Coefficient			[95% conf.	interval]
					2799328	.0056486

Number of obs = 8,063

Number of obs = 8,063

Average marginal effects

Model VCE: OIM

Expression: Pr(married), predict()

dy/dx wrt: trt_school yob sex 5.visit_month 7.visit_month

			Delta-metho	d			
	1	dy/dx	std. err.	Z	P> z	[95% conf.	interval]
trt_school		0130801	.0069457	-1.88	0.060	0266934	.0005333
yob		0680105	.0024052	-28.28	0.000	0727246	0632965
sex	1	.2311577	.0084983	27.20	0.000	.2145014	.247814
visit_month							
5		0042307	.0097508	-0.43	0.664	023342	.0148806
7		.0180713	.0090502	2.00	0.046	.0003332	.0358093

Note: dy/dx for factor levels is the discrete change from the base level.

Differnece in Differnece approach.

Having seen the previous models, it would appear the effect of the subsidies on school dropout exists both in the short term and the long-term, being slightly higher in the long term. However, it would make sense to ty to get the differentiated effect of each period. For that, a panel model is employed (see Figure 2). The results can be visualized in Table 8. According to these, it would appear that the subsidies intervencytion has a negative effect in school evasion in the short-term as well as the long-term, it being statistically significant at a level of above 5%. What this seems to imply is that the total effect of subsidies on school

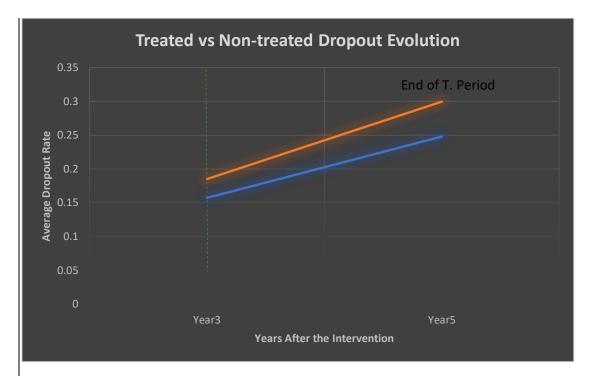
evasion, when comparado to somewone who didn't receive these is around 2.5% in the short term and around 2.7% less likelihood of dropping out of school.

Table 8

Linear regression	n			Number of o F(3, 19306) Prob > F R-squared Root MSE	bs = = = = = = = = = = = = = = = = = = =	19,310 114.39 0.0000 0.0177 .41242
dropout	Coefficient	Robust std. err.	t	P> t	[95% conf.	interval]
1.trt_school 1.post	•	.0076554	-3.64 13.35		0429031 .0976869	
trt_school#post 1 1	023748	.0118629	-2.00	0.045	0470003	0004958
_cons	.1850947	.0055426	33.40	0.000	.1742308	.1959586

Figure 2





Relationship between school marriage, children and dropouts?

Although it makes sense to asume that dropping out of school and teen marriage and pregnancy shouldn't be included into the regression due to possible bias with dropping out (both could be a product of a third variable, like irresponsability), it is possible to think that teen marriage and teen pregnancy could be relevant variables to explaining the process of dropping out of school. When estimating a regression with these two variables in Tables 9 and 10, we see that the impact of being treated decrease substantially while the explantory effect of these variables increase, especially at the short-term. This could mean that more attention should be put into the causality of these variables, and not Just assuming they are both perforando metrics.

Table 9

```
logit dropout married children trt_school if post == 0

Logistic regression

Number of obs = 4,679

LR chi2(3) = 2315.06

Prob > chi2 = 0.0000

Log likelihood = -987.52283

Pseudo R2 = 0.5396
```

dropout	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
+-	4.327641	1994044	21 70	0 000	3.936815	
•					3.792821	
trt school I	- 1336577	1260021	_1 05	0.000	3825577	1152/23
cons l	-2 825353	0928272	-30 44	0.293	-3.007291	-2 643415
Table 10						
Logistic regre	ssion				Number of obs	= 4,231
					LR chi2(3)	= 2336.73
					Prob > chi2	= 0.0000
Log likelihood	= -1273.7816				Pseudo R2	= 0.4784
	Coefficient 				[95% conf.	interval]
married	2.109351	.1208567	17.45	0.000	1.872476	2.346226
· ·					2.452414	
					4146751	
_cons	-2.577347	.0866784	-29.73	0.000	-2.747233	-2.40746
. margins, dyd:					Number of o	bs = 4,231
Model VCE: OIM Expression: Pr			1			
			1			
Expression: Pr dy/dx wrt: ma:	rried childre	en trt_schoo				
Expression: Pr dy/dx wrt: ma:	rried childre	en trt_schoo		P> z	[95% conf.	interval]
Expression: Pr dy/dx wrt: ma:	rried childre	en trt_schoo				
Expression: Pr dy/dx wrt: ma:	rried childre	en trt_schoo Delta-method std. err. .0096967	z 18.45	0.000	.1598757	.197886

Appendix

I would hace liked to repeat the main findings of the text in this sección.



Table A1

c) Possible Interaction Effects of the Subsidy (Mentioned, but placed in Appendix)

Iteration 0: log likelihood = -3616.1031
Iteration 1: log likelihood = -3256.3472
Iteration 2: log likelihood = -3222.3153
Iteration 3: log likelihood = -3222.2188
Iteration 4: log likelihood = -3222.2188

Logistic regression

Number of obs = 9,004 LR chi2(8) = 787.77 Prob > chi2 = 0.0000 Pseudo R2 = 0.1089

Log likelihood = -3222.2188

dropout	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
1.sex 1.trt_school	.8553479 1497789	.0919104	9.31 -1.54	0.000	.6752069 3407792	1.035489
sex#trt_school 1 1	 1130723 	.1300684	-0.87	0.385	3680017	.1418571
yob	 588758 	.0231685	-25.41	0.000	6341675	5433485
visit month						
_ 9	4786788	.269078	-1.78	0.075	-1.006062	.0487045
10	8135247	.2813077	-2.89	0.004	-1.364878	2621717
11	1599604	.2698813	-0.59	0.553	6889181	.3689972
12	1935362 	.3150058	-0.61	0.539	8109362	.4238637
_cons	1169.329	46.07736	25.38	0.000	1079.019	1259.639

.

Table A2

. logit dropout yob##trt_school sex i.visit_month if post == 0

Iteration 0: log likelihood = -3616.1031
Iteration 1: log likelihood = -3287.2836
Iteration 2: log likelihood = -3225.7565
Iteration 3: log likelihood = -3214.0638
Iteration 4: log likelihood = -3213.9853
Iteration 5: log likelihood = -3213.9852

Logistic regression

LR chi2(20) = 804.24Prob > chi2 = 0.0000Log likelihood = -3213.9852 Pseudo R2 = 0.1112

Number of obs = 9,004

dropout	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
yob	' 					
1986	6087783	.3355975	-1.81	0.070	-1.266537	.0489807
1987	-1.342022	.2951157	-4.55	0.000	-1.920438	7636056
1988	-1.846885	.2800821	-6.59	0.000	-2.395836	-1.297934
1989	-2.488037	.2780185	-8.95	0.000	-3.032943	-1.943131
1990	-3.067124	.2808695	-10.92	0.000	-3.617619	-2.51663
1991	-3.708385	.3010533	-12.32	0.000	-4.298439	-3.118332
1992	-3.654042	.362087	-10.09	0.000	-4.363719	-2.944364
1.trt_school	3982257	.4187955	-0.95	0.342	-1.21905	.4225984
yob#trt_school						
1986 1	.0298708	.5236579	0.06	0.955	9964797	1.056221
1987 1	.1919541	.4610569	0.42	0.677	7117008	1.095609
1988 1	.2949508	.4407094	0.67	0.503	5688237	1.158725
1989 1	.2607892	.436644	0.60	0.550	5950174	1.116596
1990 1	.028782	.442476	0.07	0.948	838455	.896019
1991 1	.10818	.4708202	0.23	0.818	8146107	1.030971
1992 1	.3134019	.5543103	0.57	0.572	7730263	1.39983
sex	.8105168	.0688728	11.77	0.000	.6755287	.945505
visit_month						
9	4962747	.2703959	-1.84	0.066	-1.026241	.0336915
10	8174601	.2824944	-2.89	0.004	-1.371139	2637813
11	1663211	.271211	-0.61	0.540	6978849	.3652427
12	2035364	.3163445	-0.64	0.520	8235603	.4164875
_cons	.784527	.3731042	2.10	0.035	.0532563	1.515798

References



Loser, J., Özler, B. and Premand, P. (2021). "What have we learned about cash transfers?". *World Bank Blogs*, Impact Development Section. https://blogs.worldbank.org/impactevaluations/what-have-we-learned-about-cash-transfers

Öztürk, E., Kardas, M., Sapmaz, A., Dönmez, U., Sahin, İ., & Çölkesen, A. (2022). The place of education in social development and its sociological foundations. Educational Research and Reviews, 17(8), 209-212.

Spiel, C., Schwartzman, S., Busemeyer, M., Cloete, N., Drori, G., Lassnigg, L., . . . Reich, R. (2018). The Contribution of Education to Social Progress*. In International Panel on Social Progress (IPSP) (Author), Rethinking Society for the 21st Century: Report of the International Panel on Social Progress (pp. 753-778). Cambridge: Cambridge University Press. doi:10.1017/9781108399661.006

Ullmann, B. Atuesta, M. Rubio and S. Cecchini, "Non-contributory cash transfers: an instrument to promote the rights and well-being of children with disabilities in Latin America and the Caribbean", Project Documents, (LC/TS.2020/154), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2021.