

Development Impact Lab Data Task

Professor Michael Kremer

Alejandro Ortiz

Table of contents

Question 1	3
Schools	3
Students	4
Conclusion	5
Question 2	5
Question 3	7

Disclaimer: The responses and content presented in this document are solely my own work. They reflect my personal understanding, analysis, and perspective. All material contained herein has been created independently by me.

Question 1

Are schools comparable across treatment groups? Are student cohorts?

To assess the comparability of the groups, I conducted four statistical tests: a simple joint significance F-test, a Wald test for robustness, individual difference-of-means tests for the continuous variables, and individual chi-squared tests for the discrete variables.

The results are as follows:

Schools

For schools I calculated the following model on the baseline data:

$$Treatment_i = \beta_0 + \gamma_{d,i} + \beta X_i + \varepsilon_i$$

Where $\gamma_{d,i}$ represents district fixed effects, and X_i are the school-level covariates including the number of: teachers, female teachers, female students, male students, number of schools in a 2 km radius, and number of latrines as well as:

- Urban dummy
- Female head teacher dummy
- Average teacher age
- The average student score

the F-statistic for this model has a value of 1.005 on 11 and 143 degrees of freedom. The corresponding p-value of this statistic is 0.445. However, because this is a Linear Probability Model (LPM), it is heteroskedastic and can yield out-of-bounds predictions. The heteroskedasticity is problematic because it can bias the standard errors, thereby rendering the F-statistic unreliable and typically increasing the probability of a Type I error. To account for this, the model is estimated using heteroskedasticity-robust standard errors. Additionally the Wald test is performed. Even though this test is asymptotically equivalent to the F-test, it serves as a robustness check on the joint significance of the model. The Wald test has a p-value of 0.166.

To test for individual significance, I conduct a series of T-tests on continuous variables. The result of these tests can be seen in Table 1. The table shows that all p-values are above 12%, with most of these p-values above 43%. The number of girls and average student score have the

lowest observed p-values (13% and 19%) which is slightly concerning. However, given that these p-values have not been corrected for multiple hypothesis testing and that they are still above the conventional 10% threshold, I do not consider this to be strong evidence of imbalance.

Table 1: School-Level Baseline Balance Check - T-test Results for Continuous Covariates

Variable	Control Mean	Treatment Mean	p-value
N. teachers	14.537	14.181	0.579
N. female teachers	7.146	7.205	0.926
N. female students	254.963	231.675	0.128
N. male students	253.573	232.952	0.171
N. schools within 2 km	1.951	2.012	0.832
Avg. teacher age	39.613	39.962	0.487
Avg. student score	249.272	255.225	0.188
N. latrines	10.949	11.627	0.433

For discrete variables I do Chi-squared tests. The results of these tests can be seen in Table 2. The lowest p-value observed is 0.37, which indicates that there is no statistical evidence to suggest imbalance across any of the variables, even at the 10% significance level.

Table 2: School-Level Baseline Balance Check - Chi-Squared Test Results for Discrete Covariates

Variable	Chi-squared statistic	p-value
Urban school	0.818	0.366
Female head teacher	0.000	1.000
District	0.004	0.952

Students

The same process is repeated for students as was applied to schools. However, because the baseline variables for students differ, the model only includes a dummy variable for female students and the year-of-birth variable as controls. The resulting F-statistic is of 1.318 on 2 and 8744 degrees of freedom. The corresponding p-value of this statistic is 0.268. The Wald test has a p-value of 0.268.

Table 3 presents the results of the t-test for the difference of means in the year-of-birth variable between the control and treatment groups. There is no statistically significant difference between these groups.

Table 3: Student-Level Baseline Balance Check - T-test Result for Year-of-Birth

Variable	Control Mean	Treatment Mean	p-value
Year-of-birth	1989.648	1989.613	0.205

Similarly, Table 4 presents the results of the randomization check for female students. The table indicates that there is no statistical evidence to suggest an imbalance between the control and treatment groups.

Table 4: Student-Level Baseline Balance Check - Chi-Squared Test Result for Female Student Dummy

Variable	Chi-squared statistic	p-value
Female student	0.58	0.446

Conclusion

All of these tests yielded p-values greater than the 10% significance level. Although a few p-values are lower, around 12%, these are infrequent. Furthermore, given that no correction for multiple hypothesis testing has been applied, these results could be due to finite sample variation. Since all tests consistently fail to indicate an imbalance between the treatment and control groups at baseline, I conclude that randomization was successful at both the school and student levels.

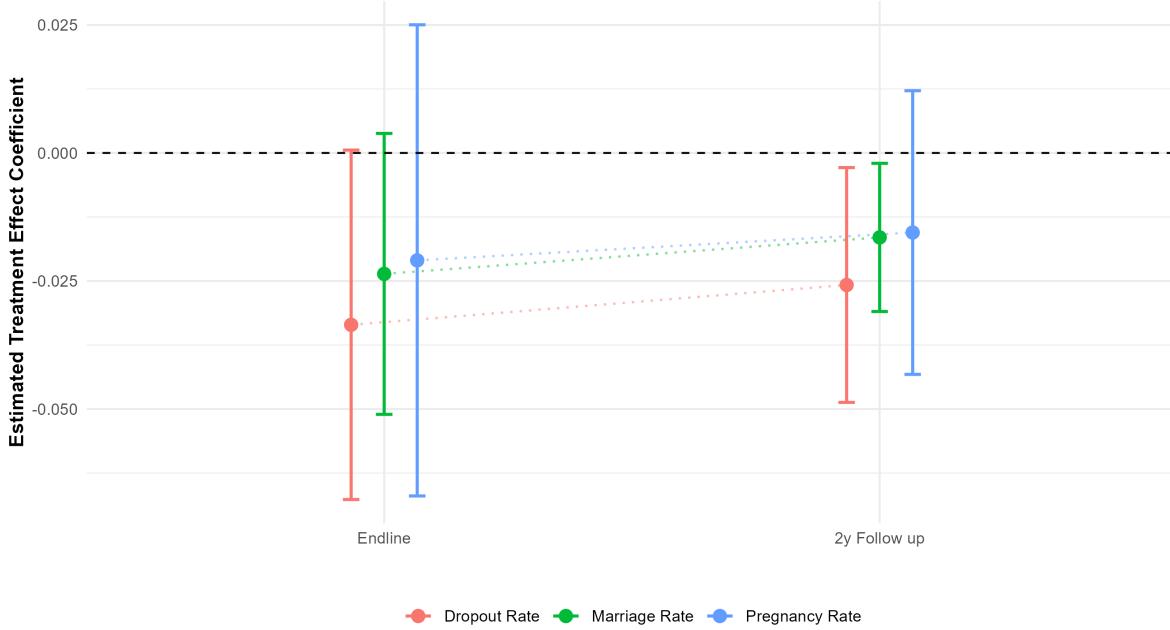
Question 2

What were the effects of the intervention on the outcomes of interest after 3 and 5 years?

Figure 1 displays the point estimates and 95% confidence intervals for the treatment effects on the main outcomes of interest at the endline (three years after the start of the intervention) and the Two-Year Follow-up (five years after the start of the intervention). The results indicate that at the 5% significance level the intervention did not have a statistically measurable effect on school evasion rates, pregnancy rates, or marriage rates after three years. However, after five years, the intervention did demonstrate a statistically significant effect on dropout and marriage rates, but not on pregnancy rates. It is important to note that since the treatment was administered at the school level, these coefficients are also presented at the school level.

Some students had missing observations in the outcome variables at the endline, the follow-up, or both. To test whether the previous results were driven by these missing data, I estimated if the treatment influenced the proportion of missing values for each outcome. The results of this

Figure 1: Treatment Effect Point Estimates and 95% Confidence Intervals on Main School-Level Outcomes



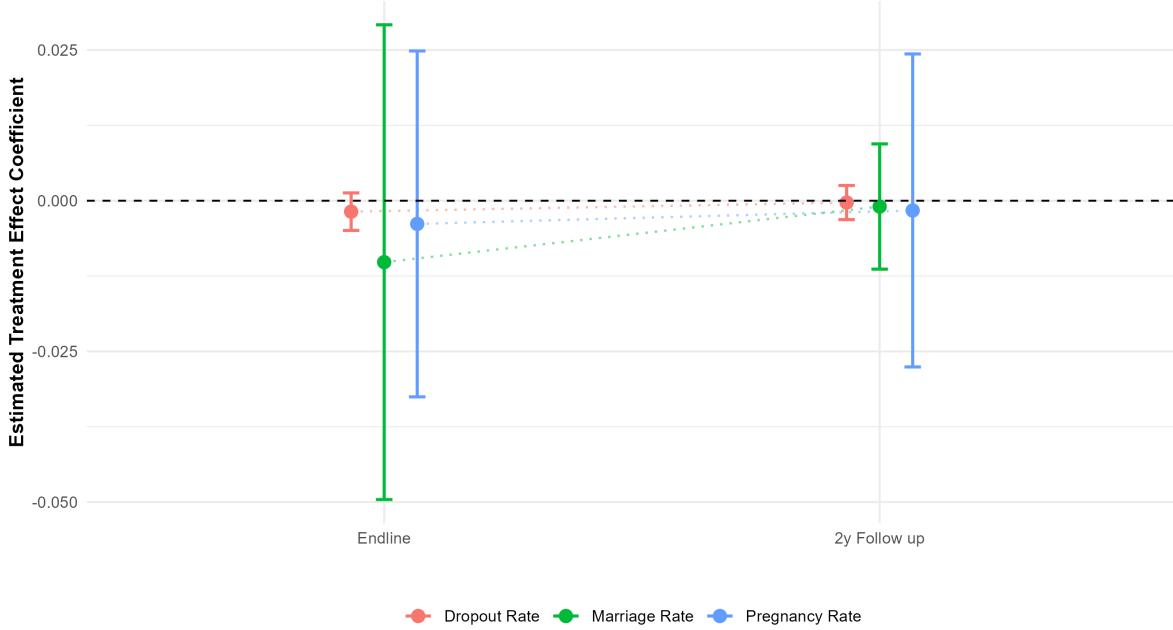
analysis can be found in Figure 2. As the plot demonstrates, there is no statistically significant effect on the proportion of missing values for any of the outcomes at either the endline or the follow-up. This indicates that missing values are not uniquely distributed between the treatment and control groups and are uncorrelated with the treatment. Therefore, the previous results were not driven by missing values.

Potential challenges to the identification of the treatment effect include possible spillover effects. For example, the program's effect in a treated school might influence students in nearby schools. This is a valid concern that could complicate the interpretation of the program's impact.

In this case, the ability to empirically test for spillover effects is limited because the only available variable indicating the connections between schools is the number of schools within a 2-kilometer radius. However, we lack information on whether these nearby schools are treated or control institutions, or their precise distance within that radius. As such, we do not possess information on the strength (distance) of the inter-school bonds or the treatment status relationship (treated and control) between them, which significantly restricts our ability to empirically evaluate spillovers.

Due to these limitations, I do not empirically evaluate the possible effect of spillovers but propose a methodology in case one sought to test for such effects. This method involves estimating a regression for each outcome of interest that includes the treatment variable, the number of schools in a 2-kilometer radius, and the interaction term between these two variables.

Figure 2: Treatment Effect Point Estimates and 95% Confidence Intervals on Missing Outcome Data Rates



The interaction term would indicate the differential effect of the number of nearby schools (potentially control schools) on treated schools and should be statistically insignificant if no spillover effects are present. Similarly, the term for the number of schools indicates the effect of having other schools (potentially treated) nearby a control school and should also be statistically insignificant if spillover effects are absent.

Question 3

Were the effects on school evasion different for girls and boys?

Table 5 presents the estimation results for the dropout dummy as a function of the treatment dummy, the female dummy, and their interaction term, for both the endline and the follow-up. The model employed is a Linear Probability Model (LPM) and is therefore heteroskedastic. All estimations were performed using heteroskedasticity-robust standard errors and were clustered at the school level to account for the correlation of outcomes among students within the same school.

As the table demonstrates, only the intercept and the coefficient for the female dummy are statistically significant at any of the conventional levels of confidence (10% or below). The coefficient indicating the differential effect, i.e. the interaction term between treatment and

female, is not statistically different from zero, which suggests that males and females were equally affected by the program. The treatment coefficient is also not statistically different from zero, which suggests that the program had no discernible effect on male students.

Table 5: Regression of Dropout on Treatment, Female Dummy, and Interaction

	Endline	2y Follow-up
(Intercept)	0.208*** (0.016)	0.108*** (0.008)
Treatment	-0.033 (0.020)	-0.008 (0.011)
Female	0.087*** (0.014)	0.052*** (0.010)
Treatment × Female	-0.024 (0.021)	-0.014 (0.014)
Num.Obs.	8717	8731
R2	0.011	0.005

* p < 0.1, ** p < 0.05, *** p < 0.01

Note:

The dependent variable is a dropout dummy. The model is a Linear Probability Model (LPM). Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the school level.