ECON 21110 Final Project

Allen (Yixin) Hu

3/13/2022

Overview of the Paper and Replication

In this replication, I will be attempting to replicate the findings and results set forth by Duflo et al. in their paper *"School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools"*. As a background, it is according to the paper that smaller class sizes (aka pupil-teacher ratios) has been a key focus on many education policy makers, while another camp of policy makers argue that these impacts without systematic reforms to education systems, teacher incentives and pedagogy is not effective in increasing pupil outcomes. In their paper, Duflo et al. presents a RCT whereby Kenyan schools were given funds to locally hire teachers with the aim of decreasing class sizes. In doing so, they are able to estimate the effects of decreased class size on student outcomes, as well as how changes in teacher incentives may affect pedagogy and these student outcomes. The authors present four main findings, of which I will be focusing on the first three for this replication exercise.

First, they find evidence suggesting that despite a reduction in class sizes by hiring new local teachers (from an average of 82 students per class to 44), there were no significant changes to student outcomes if teacher incentives and the current education system remain the same. However, they see that newly hired local teachers, when facing scrutiny and a temporary contract (i.e. new incentives) as well as their systematic difference in characteristics such as age, experience, and 'energy levels', induce a significant change in student outcomes in the positive direction. Furthermore, the authors find that empowering the community of parents by establishing a "parent's guild" within the school to hold school management accountable increases the positive outcomes for students both randomly assigned to newly hired teachers and 'pre-existing' teachers of the school in ETP schools (SBM was not established in control schools, and **SBM was only established in a subset of ETP schools**). Therefore, using these different policies, the authors argue that this new method for hiring teachers, coupled with empowering the community to hold school management accountable, may serve as a good supplement to the current education system in Kenya. While there are other significant results such as the mechanisms of improved performance, this replication will focus on these two main findings.

In a word: bringing in new young teachers into the education system (even as trainees that can become full teachers later) may benefit student outcomes. This benefit is further enhanced when the community is empowered to hold school management accountable.

The rest of the exercise will be organized as follows: In Section 2, I will present and discuss the empirical method of this paper. In section 3, I will dive deeper into the data, where it comes from, what are some key variables, and how this maps onto the author's theoretical constructs. In section 4, the most extensive section, I will present tables on the effects on student outcomes based on whether or not she was in a basic ETP (extra teachers program) school or ETP school with SBM (school based management); Effects on student outcomes based on whether or not she was randomly assigned ETP (newly contracted) teacher or TSC (originally employed) teacher within either a basic ETP school or ETP school with SBM; Finally, this section will conclude with an original extension looking at the differential treatment outcomes for girls and boys. It will be shown that girls are disproportionately disadvantaged in these treatments, whereby most of the gains are realized by the boys. Finally, in Section 5 I will conclude by describing what is learned in the paper, caveats surrounding the data and external validity, and some considerations for further research.

Section 2: Discussion on Empirical Design

Background on the Study: What is the ETP and How was it Implemented

Now the discussion turns to the background of this study. In response to a need for extra teachers to decrease class size, the International Child Support (ICS) worked with local schools to help provide additional teachers to help with teaching younger grades. This program, whereby schools were given an allowance to hire new teachers, was called the "Extra Teachers Program" (ETP for short). In terms of technicalities of language usage in this paper, all teachers hired through this program are called "ETP teachers", and schools which have been randomly chosen for this program are called "ETP schools". Of course, there are also 'pre-existing' teachers within these ETP schools. These 'pre-existing' teachers are called 'Teachers Service Commission (TSC)" teachers: effectively teachers that are centrally hired by the government. Finally, **within the ETP schools**, the 90-minute training program which empowers the community and establishes "parent's guild" to hold school management accountable is called the School-Based Management Empowerment Training (SBM), and these are only established in ETP schools.

While the ETP program only lasted 1 year, the entire study lasted for 2 years within the Bungoma region in Kenya. The authors track teacher behavior and student outcomes throughout the first year when the program is active, and also look at a follow up study to see decay in the program's effects. The dependent variables in question are an array of different measures of student welfare, such as improvement in test scores, attrition from school, and student-teacher interactions. The key independent variables are very simple in this design: it is simply indicators of whether or not a student was in an ETP school or ETP school with SBM, and whether a student is assigned to an ETP teacher or TSC teacher. The reason why this is a reasonable design is because of the initial randomization process of this study. It is also noted that 210 schools were randomly divided into a comparison group (70 schools), ETP program schools **without** SBM (70 schools), and ETP program schools **including** SBM. This essentially allows us to make the argument that because

randomization is present, there should be no systematic biases within each group that we need to explicitly account for. (The authors goes at length discussing the balance of this study, which will not be presented in this exercise due to space constraints)

Empirical Design: RCT, Implications, and Considerations

As noted briefly in our previous section, the experimental design consists of an RCT where by 210 primary schools were randomly divided into a comparison group of 70 schools, a treatment group of Extra Teachers Program (ETP) program schools, and another treatment group of ETP schools augmented by the SBM program. To note, the ETP program was specifically designed to supply more teachers to help with teaching at the lower grades, and this entire process was conducted by a non-governmental organization International Child Support (ICS), with funding from the World bank, where their original goal is to simply decrease the class size. The randomization process consists of first selecting schools in need of resources (a strong source of bias and requires considerations when arguing for external validity, as noted in our discussion section (Section 5)) and then from a randomized list alternate between ETP and non-ETP schools. Later, the school-based management program (SBM) was implemented within a randomized subset of ETP schools, whereby school committee members (parents in particular) were empowered to monitor the school's implementation of the ETP program. This training session lasted 90-min and is fairly low cost. Once programs were implemented, students would be randomly assigned to join class led by the extra teacher hired from the ETP program.

However, it is also worth noting that schools offered the opportunity to participate in the ETP and SBM programs ultimately **chose** to do so. Therefore, we will definitely observe self-selection into treatment as a part of our measurement. Moreover, it was not super clear in the paper whether the ETP program may get "passed down" to schools originally assigned to the control group, so there is a layer of ambiguity on the subjects treated in this respect. Therefore, in accord with what we learned in class, we are not actually estimating a "true" RCT, but rather at best the local average treatment effect of compilers when we are looking at the regression results.

The authors then argue that shifts in student outcomes must have been caused by the program, and not any other confounder. However, even if the authors argue that within their sample there was no significant self-selection in/out of ETP schools for students, it must be noted that the schools that are considered for this study are schools that are considered "in need of education resources", and that schools may choose to opt in/out of the study before the study begins. Moreover, the treatment group only consists of lower grades, which may all lead to strong biases if we were to extrapolate an external consistency story directly. Therefore, my main extension of this paper is a discussion on a slight qualification of the author's argument, a discussion on external validity, and suggestions of what factors and assumptions should one consider when further improving on the current analysis. All caveats discussed in this section (the compliance from schools, the bias to choose less-developed schools (resource-wise), and treatment being applied to lower grades only) will be discussed further in the discussion section (Section 5).

Section 3: Data

The data that is used to run the following results in Section 4 is collected from the open-source Harvard Dataverse Website (link here). The sample considered in this study is a cohort of approximately 13,500 children enrolled in the first grade at the end of the first school term of 2005. These students were tested in three main areas which then became our primary dependent variables of interest: **test scores on mathematics and literacy examinations**. There are also approximately 754 teachers in total that taught lessons within this sample at this time frame. Of those teachers, there were 653 TSC (centrally hired) teachers and 101 ETP (contract) teachers, these teachers were mainly measured with regards to two main criterion as well: **observed effort and pedagogical efficiency**. These will also be the main variables that we will be observing (along with others) in this paper replication.

According to the authors, five unannounced visits were made to each school by the research team, where student participation, teacher effort (as measured by in-class presence). Furthermore, the research team also confirmed in these visits that the students did not cross over and go to different schools or different classes. This helps alleviate the main concern with compliance in the experimental design. By confirming the consistency of treatment, we can effectively put the concern of attrition of compliance to rest, which further strengthens the causal argument of this paper. In all, approximately 94% of all students were found in their assigned section, and the only anomaly which occurred when class sections were combined happened less than 11% of the time.

Regarding test scores, the math and literacy scores were obtained via a standardized question which ranged from identifying numbers and letters to subtracting two-digit numbers. It is also worth noting that these tests were administrated and graded blindly to avoid any confounding factors of favoritism or otherwise. These tests were administered both at the end of the program and the one year follow up, which allows a measurement of decay. The way in which these tests were administrated does not leave out a lot of room for extra concerns regard the testing itself.

As a confirmation of the distribution of test scores, a plot which mimics that of the one presented in the paper is given below along with a plot of the prior scores. In particular, focus should be drown to the two distributions "Basic ETP" and "ETP with SBM". We see that in the prior period, the two groups have very similar distributions of scores. However, in the posterior period where the program had just ended, we see graphically that the school where SBM was also implemented has some advantages in outcome relative to both the control group as well as the "Basic ETP" group. This is all to show that show that there should be some effects of the program, both normal and SBM-augmented, for student outcomes.

Beyond the school level, we can also check the distributions with respect to the teachers level. Now with a more granular view, we can see how the 90-min SBM training program may have a significant effect on individual outcomes. Notice that apart from the control group that is slightly systematic different, as long as the student is from a ETP school, their prior score distributions look very similar. However, when we move to the post period, we see how these scores begin to differentiate. What is most interesting is how

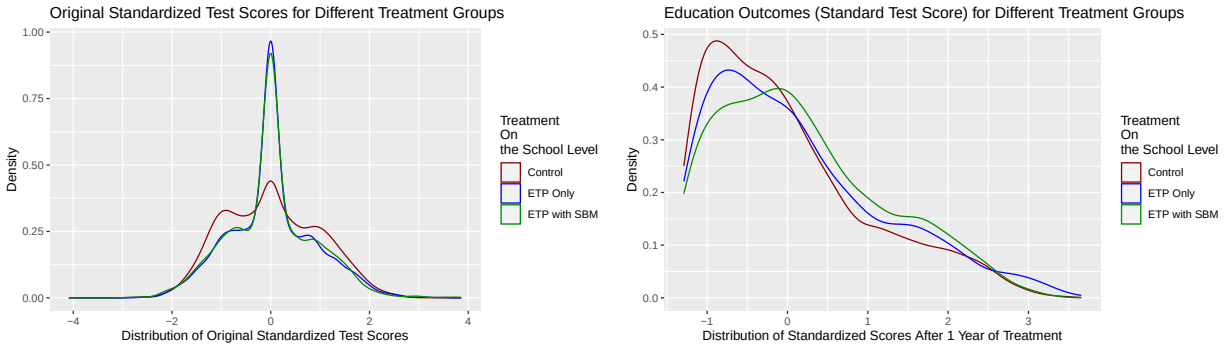

Figure 1: Comparison between the pre and post period test scores based on different school-level treatments

the SBM program actually improves student outcomes for both TSC teachers as well as ETP teachers. This may hint at the major benefits for a relatively cheap intervention which empowers parents to participate in school management.

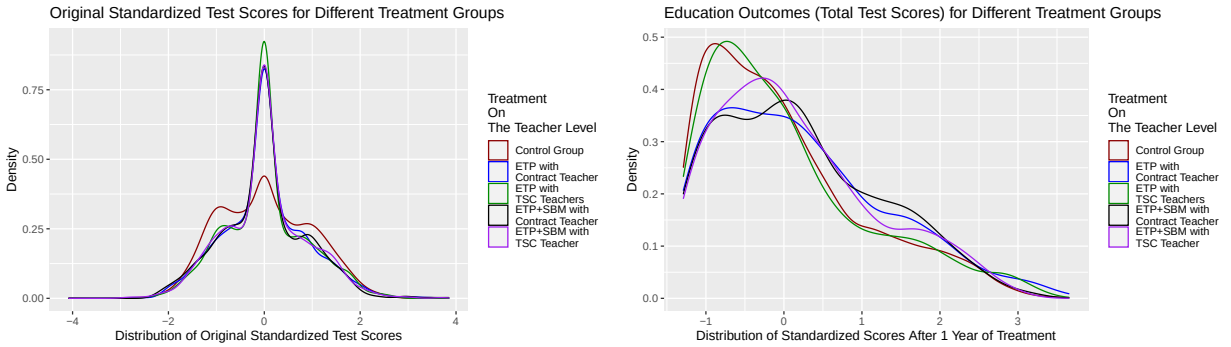

Figure 2: Comparison between the pre and post period test scores based on different teacher-level treatments

This concludes the discussion on data source, collection, and visualization of intuitive results. Armed with a direction, the discussion now turns to the regression analysis that has been proposed by the authors. In the following section (Section 4), I specifically will dive into these preliminary findings more concretely with regression analysis to see to what extent do these positive effect hold.

Section 4: Empirical Results of Student's Measure (Class Participation, Attrition, and Test Outcomes)

School Level Treatment Analysis

In this section, I will briefly go over the empirical results and how to interpret said results. Tables 1 and 2 correspond to the eight "Panel A" columns in Table 2 in the original paper, whereas tables 3 and 4 corresponds to the eight "Panel B" columns in Table 2 in the original paper. Here, we are specifically looking at the effects of enrolling in a ETP or ETP with SBM school on student's class participation rate, attrition (i.e. dropping out of school), and of course our main variables: education outcomes as measured

by the mathematics and literacy test scores. The exact regression follows the one indicated by the paper, where:

$$Y_{ij} = \beta_0 + \beta_1 Basic\_ETP_{ij} + \beta_2 ETP\_w\_SBM_{ij} + \sum_{k=3}^{n} \beta_k X_{ij} + U_{ij}$$

Where Y_{ij} is the outcome variable of interest (in our specification, class participation, attrition, and test scores). $Basic\_ETP_{ij}$ is a dummy variable which equals to 1 if the student i was attending a school j with the basic ETP program. $ETP\_w\_SBM_{ij}$ is a dummy variable which equals to 1 if the student i was attending a school j with the ETP program plus the 90-minute school-based management program (SBM). X_{ij} is a vector of control variables, which (as the paper shows) does not significantly alter the results. This means that the results are robust up to randomization, and further controls are not necessarily needed. This then can imply that the randomization was indeed a successful process. (That being said, we do see significant interaction terms on gender, more on this in the extension section)

Regarding the β coefficients, the interpretations are as the following: β_0 represents the outcome variable of the control group. For example, β_0 when testing for percentage of class attended represents the percentage of class attended for students situated in a control group school. Thus in the same example, β_1 is interpreted as the effect of going to a ETP school on class participation, and then β_2 is the effect of attending a school where by the ETP program was augmented by the SBM training.

There are many things to consider when reading this analysis. As a quick recall, note that there are two layers of randomization: schools were randomly assigned to programs, and within schools students were randomly assigned to teachers. However, as noted in our previous discussion on the empirical method, the following β values presented are to be interpreted to be the LATE (local average treatment effect) at best, since we know from Section 3 that the compliance was not 100% (where the estimated compliance is about 94%). Furthermore, because there was some degree of self-selection (where schools can still have a say in whether or not to participate) and also some degree of selection bias (only 'needy' schools were selected as potential observations), this further requires a caveat in which we cannot directly say that this result can be interpreted as an ATE, and also limits the potential generalizations (external validity) of the argument. Furthermore, attrition also needs to be taken into consideration (which is actually a variable we are measuring), therefore, I will follow the descriptions of the paper to reweigh endline test scores to take attrition into account (Note: Notice that I have retrieved a different estimate than the paper, this is almost certainly due to inconsistencies between the way in which I implement the reweighing mechanism.)

Now that a basic intuition of the econometric equation is established the regression results are presented below:

Table 1:

	Dependent variable:	
	Percentage Class Attended	Attrition at Endline Test
	(1)	(2)
basic_ETP	0.002	−0.015
	(0.004)	(0.011)
ETPwSBM	0.020***	−0.031***
	(0.004)	(0.011)
Constant	0.874***	0.200***
	(0.003)	(0.006)
Observations	51,656	8,012
R^2	0.007	0.001
Adjusted R^2	0.007	0.001
Residual Std. Error	0.348 (df = 51651)	0.387 (df = 8008)
F Statistic	96.411*** (df = 4; 51651)	3.940*** (df = 3; 8008)
Note:		*p<0.1; **p<0.05; ***p<0.01
	Control variable vector is omitted in the visualization only to save space.	

Notice that discounting the last 3 columns whereby the rescaling was not explicitly given, the point estimates of these tables 1 and 2 give the same results as the paper. On this note, because I can only go off of limited information, I attempted to resale the scores as close to the instructions given by the paper as possible. However, I was not able to arrive at the same values. This is almost certainly due to the fact that the reweighing scheme I have used is not identical to that used in the paper. However, because the exact formulation (which is akin to a kernel function as referenced by the paper) is not explicitly given, I simply cannot replicate the final 3 columns effectively. That being said however, we see that the **relative** sizes and the signs of these estimates agree with those provided by the paper. Moreover, if we were to 'recenter' the distributions of βs by interpreting all beta values as $\beta_0 + \beta_i \forall i$, then we are able to retrieve similar values as the paper.

Moving onto the confidence intervals, I am able to retrieve similar confidence intervals for the first two columns. However, when moving into the test scores in table 2, the scores explicitly reported in the tables are different from those presented in the original paper. I believe that this is almost certainly due to a standard error clustering issue, whereby we are essentially making assumptions about the standard errors which in reality violate MLR5. Therefore, I use a **robust clustering in standard errors** and I am able to retrieve the values presented in the paper by clustering on the school ID level (see code for the exact implementation, where the values are given right after each regression code). Furthermore, this hypothesis is also supported by the original paper, as the authors also note that the teaching quality between schools are very different, and thus it would be unfair to lump all of these standard errors together and evaluate them accordingly.

Now moving onto the actual interpretations of the results. In table 1, we see that the basic ETP program does not significantly affect the class participation rate nor does it help significantly with regards to the attrition at the endline test. However, we see that just like the original paper, the 90-min SBM program is able to significantly increase participation and decrease attrition. Since the SBM program is highly cost-efficient, this hints at very high benefits for a little cost.

Looking at table 2 below, note that all schools are able to significantly improve the test scores in math and literacy. In all, I find that the basic ETP program (effectively decreasing class size) does not actually contribute to higher educational outcomes, rather it is when the community is empowered to hold school management accountable that positive spoils are observed. Furthermore, once we 'recenter' the estimates in column 4, 5, and 6 where by the intercept is zero (as expected by the standardization of mean 0 variance 1) we are able to get something extremely close to the original estimates in the paper. All of this is to further enhance the argument that reductions in class size is not necessarily effective if other factors such as community empowerment is also in place.

Table 2:

	Dependent variable:					
	Total Score	Math Score	Literacy Score	Weighted Total Score	Weighted Math Score	Weighted Literacy Score
	(1)	(2)	(3)	(4)	(5)	(6)
basic_ETP	0.142***	0.133***	0.123***	0.509***	0.469***	0.443***
	(0.031)	(0.030)	(0.031)	(0.053)	(0.050)	(0.056)
ETPwSBM	0.196***	0.214***	0.141***	0.574***	0.560***	0.472***
	(0.031)	(0.031)	(0.032)	(0.055)	(0.051)	(0.058)
Constant	0.045**	0.038**	0.043**	−0.360***	−0.301***	−0.340***
	(0.019)	(0.019)	(0.019)	(0.031)	(0.029)	(0.033)
Observations	6,533	6,533	6,536	5,914	5,914	5,917
R^2	0.014	0.013	0.010	0.026	0.027	0.017
Adjusted R^2	0.013	0.012	0.009	0.025	0.026	0.016
Residual Std. Error	1.021 (df = 6529)	1.006 (df = 6529)	1.039 (df = 6532)	1.649 (df = 5910)	1.546 (df = 5910)	1.741 (df = 5913)
F Statistic	29.822*** (df = 3; 6529)	27.791*** (df = 3; 6529)	21.433*** (df = 3; 6532)	52.514*** (df = 3; 5910)	54.604*** (df = 3; 5910)	33.538*** (df = 3; 5913)

Note: *p<0.1; **p<0.05; ***p<0.01
Control variable vector is omitted in the visualization only to save space.

Teacher Level Treatment Analysis

Now to look more granularity at the teacher level effects. At the teacher level, the econometric specification becomes the equation below. In this formulation, each of the betas can be interpreted similarly to the betas in the previous econometric expression, whereby the beta is the effect of being assigned a particular type of teacher in a particular treatment group. Now with the formulation out of the way, the discussion turns to the numerical results.

$$Y_{ij} = \beta_0 + \beta_1 BasicETP\_TSC_{ij} + \beta_2 BasicETP\_Contract_{ij} +$$
$$\beta_3 ETP\_SBM\_TSC_{ij} + \beta_4 ETP\_SBM\_Contract_{ij} + \sum_{k=5}^{n} \beta_k X_{ij} + U_{ij}$$

Looking at table 3, we see that I am able to get very similar results to that of the original paper (again, the standard errors here requires clustering which will be given in the code). Here, We see that even in the point estimates, we have a very interesting result. In particular, notice that overall, basic ETP does not have a significant effect for either types of teachers. However, once SBM is implemented, there is a statistically significant effect for contract teachers, who are younger, more active, and facing a limited contract. This, the authors argue, could serve as a motivation to implement these temporary teachers in place before they are "tenured" by the TSC. This could have positive effects in both student engagement and student attrition.

Table 3:

	Dependent variable:	
	Percentage Class Attended	Attrition at Endline Test
	(1)	(2)
basic_ETP_tsc	−0.007	−0.001
	(0.005)	(0.013)
basic_ETP_contract	0.011**	−0.031**
	(0.005)	(0.014)
ETPwSBM_tsc	0.015***	−0.023*
	(0.005)	(0.014)
ETPwSBM_contract	0.025***	−0.039***
	(0.005)	(0.014)
Constant	0.874***	0.200***
	(0.003)	(0.006)
Observations	51,656	8,009
R^2	0.008	0.002
Adjusted R^2	0.008	0.001
Residual Std. Error	0.348 (df = 51649)	0.387 (df = 8003)
F Statistic	66.155*** (df = 6; 51649)	3.100*** (df = 5; 8003)
Note:		*p<0.1; **p<0.05; ***p<0.01

Control variable vector is omitted in the visualization only to save space.

Looking at table 4, we find basically the same story but stronger: we see that the only significant terms are when the teacher assigned to a student is on a contract. This further enhances the idea that a pure decrease in class size is not necessarily a good approach to increase student performance and engagement. Rather a shift in teacher's incentives and community empowerment is also necessary.

Table 4:

	Dependent variable:					
	Total Score	Math Score	Literacy Score	Weighted Total Score	Weighted Math Score	Weighted Literacy Score
	(1)	(2)	(3)	(4)	(5)	(6)
basic_ETP_tsc	0.043	0.012	0.063	0.410***	0.349***	0.382***
	(0.040)	(0.039)	(0.040)	(0.071)	(0.066)	(0.075)
basic_ETP_contract	0.243***	0.256***	0.185***	0.628***	0.603***	0.525***
	(0.040)	(0.039)	(0.040)	(0.069)	(0.065)	(0.073)
ETPwSBM_tsc	0.175***	0.200***	0.119***	0.522***	0.524***	0.417***
	(0.040)	(0.039)	(0.041)	(0.071)	(0.067)	(0.075)
ETPwSBM_contract	0.219***	0.231***	0.165***	0.644***	0.612***	0.544***
	(0.041)	(0.041)	(0.042)	(0.073)	(0.068)	(0.077)
Constant	0.045**	0.037**	0.043**	−0.361***	−0.302***	−0.342***
	(0.019)	(0.019)	(0.019)	(0.031)	(0.029)	(0.033)
Observations	6,531	6,531	6,534	5,912	5,912	5,915
R^2	0.016	0.016	0.011	0.028	0.029	0.018
Adjusted R^2	0.015	0.016	0.010	0.027	0.028	0.017
Residual Std. Error	1.020 (df = 6525)	1.005 (df = 6525)	1.038 (df = 6528)	1.656 (df = 5906)	1.551 (df = 5906)	1.751 (df = 5909)
F Statistic	21.321*** (df = 5; 6525)	21.723*** (df = 5; 6525)	14.217*** (df = 5; 6528)	33.882*** (df = 5; 5906)	35.650*** (df = 5; 5906)	21.512*** (df = 5; 5909)

Note: *p<0.1; **p<0.05; ***p<0.01

Control variable vector is omitted in the visualization only to save space.

Extention on the Differential Outcomes between Girls and Boys

Here, I will aim to see whether there is still differential impact between the ETP, ETP with SBM, and Control groups especially for girls in the program. As noted by the paper, Kenya can be (overarchingly) understood as a patriotically society, whereby females usually face significant barriers to opportunities such as education and employment. The economic interest in this question lies in the program's long-run cost effectiveness when strengthening educational outcomes and equality in said outcome. In the paper, the authors checked to see that the outcomes did not vary with respect to the "girl" variable when considered independently, however, they did not necessarily go at length about the differential impact of this program on girls. Therefore, I will be effectively interacting the girl variable with our previous independent variables to see if there is a significant difference.

Now moving from the rationale into the regression, the specification is very simple, where we take our original expression in the previous section and attach interaction terms to each variable. For the school-level effects, we have

$$Y_{ij} = \beta_0 + \beta_1 Basic\_ETP_{ij} + \beta_2 ETP\_w\_SBM_{ij} + \beta_3 Girl_{ij}$$
$$+\beta_4 Basic\_ETP_{ij} \cdot Girl_{ij} + \beta_5 ETP\_w\_SBM_{ij} \cdot Girl_{ij} + \sum_{k=6}^{n} \beta_k X_{ij} + U_{ij}$$

The interpretation of the terms $\beta_i \forall i \in \{0, 1, 2, 6...\}$ is the same as previously, where as β_3 represents the independent effect of the student being a girl on the outcome variable. The two most interesting βs will be β_4 and β_5, whereby these represent the differential impact of these programs on girls relative to boys.

A similar regression specification is made for the teacher-level effects. Effectively interacting the variable 'girl' on every teacher-level dummy that was provided. Explicitly, we have:

$$\begin{aligned} Y_{ij} = \beta_0 &+ \beta_1 BasicETP\_TSC + \beta_2 BasicETP\_Contract + \\ &\beta_3 ETP\_SBM\_TSC + \beta_4 ETP\_SBM\_Contract + \beta_5 Girl + \\ &\beta_6 Girl \cdot BasicETP\_TSC + \beta_7 Girl \cdot BasicETP\_Contract + \\ &\beta_8 Girl \cdot ETP\_SBM\_TSC + \beta_9 Girl \cdot ETP\_SBM\_Contract + \sum_{k=5}^{n} \beta_k X_{ij} + U_{ij} \end{aligned}$$

In the tables presented (5 and 6), we see a very significant result in the participation and attrition variables. In particular it is worth noting that the participation rate for class time is significantly lower for girls in both treatment groups (ETP and ETP with SBM). Looking first specifically at the participation rate, we see that the effect on participation rate specific to girls from the two treatment arms are significantly negative, and once considered with the estimator for the treatment, the treatment effect on girls becomes insignificant (for example, the point estimate for girls who participated in the ETP program has an effective ($\beta_1 + \beta_3 + \beta_4 = 0.010 + 0.000 + -0.017 = -0.007 \pm 0.017$) change in participation. Notice then that this figure

Table 5:

	Dependent variable:	
	Percentage Class Attended	Attrition at Endline Test
	(1)	(2)
basic_ETP	0.010^{*}	-0.021
	(0.005)	(0.015)
girl	0.0003	0.028^{**}
	(0.004)	(0.012)
ETPwSBM	0.028^{***}	-0.037^{**}
	(0.005)	(0.015)
basic_ETP:girl	-0.017^{**}	0.012
	(0.007)	(0.021)
girl:ETPwSBM	-0.015^{**}	0.014
	(0.008)	(0.022)
Constant	0.870^{***}	0.186^{***}
	(0.003)	(0.009)
Observations	51,656	8,012
R^2	0.008	0.003
Adjusted R^2	0.007	0.003
Residual Std. Error	0.348 (df = 51649)	0.387 (df = 8005)
F Statistic	65.429^{***} (df = 6; 51649)	4.628^{***} (df = 6; 8005)
Note:		*p<0.1; **p<0.05; ***p<0.01

Control variable vector is omitted in the visualization only to save space.

is insignificant at the 95% level and implies that the increase in participation was not evenly distributed between boys and girls. In particular, girls saw effectively no change in in-class participation. More over, we see from the second column that no treatment was effective in decreasing the attrition rate in girls. This means that if the program was to be extended in this context for the sake of equality of educational outcomes, this result might hint at its ineffectiveness in brining forth said results.

Now turning to a more granular analysis, I will present the effects and outcomes at the teacher-level.

There are a few interesting results from this regression. The table indictates that while all teachers except the TSC teachers in the basic ETP group had a positive effect on attrition rate, we see that the effects are nearlly completely nullified when the student is a girl. Further, we see that these effects on attrition does not differ with respect to gender. However, what is quite puzzling and interesting is the first column. In particular, notice that the interaction terms between girl and "Basic ETP with a TSC teacher" and "ETP and SBM with a ETP teacher" is statistically null. This means that most of the disproportionate effects that we have observed in the previous section on school level effects comes almost entirely from ETP teachers in the basic ETP group, and TSC teachers from the SBM group. However, this goes against the grain of the main predictions of the paper, whereby we should observe a level improvement when the SBM program is implemented. One explaination that I can possibly make is through one observation made in the original paper: it is noted that a large proportion of contract teachers are male relative to the TSC teachers. Therefore in the basic ETP case, female students might not be as inspired by these contractual teachers (It would be ideal to have data on whether or not the teacher was female, but it was not provided in this dataset, which gives room for future papers to improve upon). However, this line of logic cannot

Table 6:

	Dependent variable:	
	Percentage Class Attended	Attrition at Endline Test
	(1)	(2)
basic_ETP_tsc	−0.004	0.002
	(0.007)	(0.019)
girl	0.0003	0.028**
	(0.004)	(0.012)
basic_ETP_contract	0.025***	−0.048**
	(0.007)	(0.019)
ETPwSBM_tsc	0.026***	−0.038*
	(0.007)	(0.019)
ETPwSBM_contract	0.029***	−0.037*
	(0.007)	(0.020)
basic_ETP_tsc:girl	−0.006	−0.006
	(0.009)	(0.027)
girl:basic_ETP_contract	−0.029***	0.034
	(0.010)	(0.027)
girl:ETPwSBM_tsc	−0.022**	0.029
	(0.010)	(0.027)
girl:ETPwSBM_contract	−0.008	−0.003
	(0.010)	(0.029)
Constant	0.870***	0.186***
	(0.003)	(0.009)
Observations	51,656	8,009
R^2	0.008	0.004
Adjusted R^2	0.008	0.003
Residual Std. Error	0.347 (df = 51645)	0.387 (df = 7998)
F Statistic	40.858*** (df = 10; 51645)	3.399*** (df = 10; 7998)

Note: *p<0.1; **p<0.05; ***p<0.01
Control variable vector is omitted in the visualization only to save space.

explain the results in the SBM case, where we see a flip, and TSC teachers in the SBM case are the ones that are bringing down the class participation rate for girls to a statistically zero number. With SBM being implemented, teachers should be more accountable and thus be encouraged to interact and take care of their students, yet this is not the case, whereby TSC teachers are the ones bringing down involvement where they were not in the basic ETP schools. Nevertheless, this "flip in results" when the intervention should be a level improvement presents a puzzle that can and should be addressed in future studies. As a quick aside on scores, we do not see a lot of differential effects of the treatments with respect to gender (note, that columns 4, 5, and 6 are scaled based on paper descriptions and are not necessarily in alignment with how the paper truly scaled them).

Section 5: Conclusion

In all, this replication of the paper by Duflo et al. indicates that a sheer decrease in classroom size might not be a sufficient policy to improve student welfare. Here, it is noted that beyond a simple decrease in class size, other programs such as community empowerment (SBM) and switching up the teacher's own incentives (limited-time contracts renewed based on performace) might also be key in improving student's outcomes. It should be noted again that these results are at best interpreted as LATEs, whereby there was a degree of bias in the selection process (schools were generally less wealthy) and a slight degree of self selection (schools ultimately had a choice of opting in). Beyond this, it is also worth noting that this experiment is conducted almost solely on a generally more low income and early elementary school group. This poses a lot of questions with regards to generalization. For example, it is also noted by the authors that these contract teachers are often times less experienced. Would the same positive effect be seen if the cohort was in an older grade when more rigorous academic effort is needed? This question is a small one, but it goes to show that while the results definitely interesting and very well thought through, they are not necessarily one-to-one scalable to other scenarios where the student are of a different grade, the social norms are different, and the average income level is different. That being said, this paper still sheds light as to what are some of the main components to drive up student welfare in early childhood education, specifically the need to not only decrease class size, but also consider empowering the community, bring new incentives to teachers, and as my original extension shows, a careful consideration of gender disparities in educational outcomes.

References (Original Paper):

Duflo, E., Dupas, P., & Kremer, M. (2019, November 13). *School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan Primary Schools.* Harvard Dataverse. Retrieved March 17, 2022, from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2F9534YA