Project Title: Evaluating the Impact on Low-income Children and Families of Access to a Private Comprehensive Schooling Model: Experimental Evidence from Mexico

Project Description

This evaluation will provide evidence on the impact of a comprehensive private schooling alternative in Mexico City on the outcomes of poor students. Outcomes include student achievement (learning outcomes, as measured by test scores), parent involvement, and student behavioral outcomes. In addition, the evaluation will provide qualitative evidence on the implementation of these models, including costs of the implementation. The evaluation will focus on the CHM school model in Mexico. Mexico is an interesting example because it has poor populations that have been underserved by traditional public school models.

The results of the evaluation, while informative about the impact and cost-effectiveness of this particular schooling model, should be viewed with caution if the ultimate goal is scalability and large-scale replication. This is an evaluation of a single school in Mexico City that has high input requirements and a particular funding mix. As such, this evaluation should be seen as a proof of concept.

This proof of concept, however, could serve as a stepping-stone towards a broader evidence base of what works and what doesn't in terms of alternative schooling models for the poor. In Latin American countries such as Mexico or Colombia, for instance, there are severe capacity constraints in the public schooling system that disproportionately affect poor children. In Mexico City, for instance, most public schools—the only real alternative to low-income children—have switched to double, and in some cases triple shifts because the available infrastructure cannot accommodate growing demand. Research indicates that schools' double-shifts are the main obstacle to providing full-day education in Mexico, and that students in the afternoon shifts receive lower-quality education and have lower achievement. Similarly, in Colombia, the government has, for many years now, contracted out enrollment expansions with the private sector (via vouchers for private schooling or direct contracts) or through public-private partnerships (via concession schools) because public schools are overcrowded. So proof of concept of the cost-effectiveness of a comprehensive low-cost private school is a valuable first step in understanding whether in contexts such as Mexico, where there are few high-quality affordable private school alternatives for low-income children and no charter school movement to speak of, there are promising practices that should be considered if such a movement is to become part of the education policy agenda.

Second, there are successful examples in the U.S. of the expansion of similar "No excuses" private schools. KIPP is one of them. When KIPP began in the U.S. in the 1990s, the national charter school movement was just beginning. KIPP's founders began with a special fifth grade program for about 50 students in which they expanded the school day to 9½ hours, added Saturday classes, extended the school year into the summer months. They insisted on good behavior and required students to call them at home if they had any questions about their homework. The founder's premise was to let academic results of their effort speak for themselves. Today, there are more than 4,000 charter schools in the U.S. KIPP alone has 57 schools in 17 states and the District of Columbia and has produced the greatest achievement gains ever seen in so large a private school network. More than 80 percent of its students are low-income, and 95 percent are black or Hispanic. It obviously requires great managerial skills and funding, but in some sense KIPP started in the U.S. in the 1990s at a much more basic level than Christel House in Mexico in the 2000s.

Third, it is not clear that the parents who apply to this school do so because they only care about test-scores. In fact, a recent survey conducted by the Instituto de Fomento e Investigacion Educativa (IFIE) and the National Science Council in Mexico, found than only one out of every four parents in Mexico City's metropolitan area was even aware of their students' test scores in ENLACE.1 From what we know and saw during our field-visit during the proposal-writing stage, there are other dimensions that seem equally if not more important to parents. These include the length of the school day (many parents work long hours), the nutritional supplementation and psycho-pedagogical support (many children begin with nutritional and cognitive deficiencies) and the school for parents. Moreover, at Christel House, new-student admissions take place mostly in first grade. In this sense, the comment about the school only impacting children of parents who care about test-scores, while valid in Charlotte-Mecklenburg, is less of a concern for this study.

Project objectives include:

1. To document Early Grade Language and Math Impacts on Students
The first evaluation objective will be to document using a rigorous experimental research design how a comprehensive private schooling alternative affects early-grade student outcomes among disadvantaged populations in Mexico City. To measure these outcomes we will use the Early Grade Reading Assessment (EGRA) and the Early Grade Math Assessment (EGMA). These assessments have been psychometrically validated for the student population that is the target of this evaluation and have been used in many other countries in Latin America and Africa to assess impacts of educational programs that target basic literacy and numeracy during the first school years. In Year 2 we will report one-year impacts for the 2013 1st grade lottery cohort, the first of two cohorts that we will employ in our analyses (see below).

2. To Document Costs
The third evaluation objective will be to measure CHM costs in relation to those in counterfactual public school conditions. To do so we will use a method to consistently identify value and distribute over time all costs, including costs of personnel, facilities, equipment and materials, other school inputs and client inputs.

The evaluation design will be **experimental.** Every year, there are more eligible applicants than available first-grade slots at Christel House Mexico (CHM). CHM has

[1] http://www.ifie.edu.mx/web/?page_id=29

determined that, as of 2013, a lottery assignment mechanism is the fairest way to allocate oversubscribed first-grade positions among eligible applicants (see enclosed letter supporting the randomized assignment). For the 2013 and 2014 applicant lotteries we will—in close coordination with CHM—oversee the lottery implementation to ensure it is valid.[2]

Our evaluation design will take advantage of the lottery assignment for first-grade positions for the 2013 and 2014 applicant cohorts. The treatment/control allocation rule will, therefore, be a publicly held lottery in June 2013 for the 2013 first-grade preselected applicants that do not have sibling priority. Similarly, a lottery among first-grade applicants in the 2014 cohort that do not have sibling/orphan priority will randomly allocate students to treatment and control conditions.

Because the lottery will allocate applicants to treatment and control groups, the assignment probabilities are independent of applicant's potential outcomes, whether these are observed or missing. Moreover, because the number of treated units is fixed and predetermined by CHM's capacity constraints and sibling/orphan priorities, the allocation rule will be a *completely randomized experiment* (Imbens and Rubin, 2007).

In each of the two cohorts of our evaluation sample (2013 and 2014), the evaluation will include two arms: one treatment and one control arm. For each applicant cohort, in the treatment arm applicants without sibling priority are randomly assigned to a first-grade slot at CHM. Applicants without sibling priority who do not gain admission to a first-grade slot at CHM are assigned to the control condition. The unit of assignment is the applicant, which implies that there are no assignment clusters.

Because the allocation rule is a completely randomized experiment, we expect that at baseline there will not be systematic differences in the distribution of observable and unobservable characteristics of applicants assigned to the treatment condition and those assigned to the control condition. Therefore, the lottery creates the gold-standard counterfactual because, on average, applicants in the control condition will represent what would have happened to treatment applicants in the absence of the treatment. CHM has worked in the same neighborhood for the past six years and they know that most non-admitted applicants attend public neighborhood schools in the catchment area. Therefore, we are able to hypothesize that the counterfactual condition for treated applicants will likely entail the experience of attending one of the elementary schools in CHM's student catchment areas, mainly the Alvaro Obregon neighborhood. For reference, relative to other public schools in Mexico City, public schools in Alvaro Obregon are relatively low performing, ranking tenth among Mexico City's 16

[2] One potential concern in this context that relates to the lottery validity is how to allocate slots of lottery winners who opt not to enroll. In recent past cohorts, only two applicants who gained acceptance to CHM have opted not to enroll, which suggests that this concern is minor. However, to make sure we preserve the integrity of the lottery assignment, as part of the lottery we will also draw a (randomly chosen) short waitlist group. When an opt-out occurs, we will offer a slot to the next-in-line applicant from the waitlist and record this applicant's lottery status accordingly. We will carry out this process until all slots are filled up.

neighborhoods. Student-teacher ratios among Alvaro Obregon public elementary schools are also slightly larger than then citywide average.

Randomization implies that we can estimate the impact of being offered a first-grade slot at CHM—the Intent-to-Treat (ITT) effect—by a simple difference in mean outcomes between treatment and control groups. Although we will report mean outcome differences between the two groups, we will also report regression-adjusted ITT effects in which we control for baseline outcomes in the cognitive and socio-emotional test that CHM administers. Regression-adjusted ITT effects have, in this particular setting, considerably more statistical power because CHM's baseline cognitive and socio-emotional tests are strong predictors of early-grade—1st grade through 3rd grade—outcomes, as we indicate in the Power Analysis section below.

As Table 1 indicates, as part of this evaluation we will collect outcome data in two post-lottery years (2014 and 2015) for the 2013 applicant cohort and in one post-lottery year (2015) for the 2014 applicant cohort. Given that for this evaluation we have a fixed cross-sectional sample in each applicant cohort, the highest statistical power gains are achieved by having one baseline and two post-treatment data collection rounds (McKenzie, 2012).

Table 1. Data collection years for the applicant cohorts in the evaluation sample

	Which data is collected in year…?			
	FY 2013	FY 2014	FY 2015	FY 2016
2013 Lottery Applicant Cohort (1st Cohort)	Baseline 1st Cohort	One-year follow-up 1st Cohort	Two-year follow-up 1st Cohort	None
2014 Lottery Applicant Cohort (2nd Cohort)	None	Baseline 2nd Cohort	One-year follow-up 2nd Cohort	Two-year follow-up 2nd Cohort

To maximize statistical power, we propose to report estimates that pool data from the 2013 and 2014 applicant cohorts. Pooling two equally-sized lottery cohorts—as is the case in CHM's application process—would reduce the estimated standard errors by about $1/\sqrt{2}$, or a 30 percent reduction (Wooldridge, 2001). Because lottery conditions and/or applicant pool characteristics may vary by year, we also will disaggregate results by cohort but we will have more limited power to detect cohort-specific effects.

To estimate the regression-adjusted ITT effects of being offered a position at CHM when we pool all post-treatment outcomes data for the two applicant cohorts (i.e. after the 2015 data collection year) we will estimate ANCOVA models that have the form:

(1) $$Y_{ict} = \alpha_c + \gamma_t + \phi\, Z_{ic} + X_{ic}'\beta + \epsilon_{ict}$$
$$Y_{icgs} = \phi\, Z_{icgs} + X_{icgs}'\beta + \delta_{cgs} + \epsilon_{icgs}$$

where Y_{ict} is an outcome for applicant i in applicant cohort c (2013, 2014), measured in post-treatment year t (2014, 2015); α_c and γ_t are cohort and year-of-measurement fixed effects; Z_{ic} is an indicator that takes the value of one if applicant i is awarded a first-grade slot through the lottery and zero otherwise; X_{ic} is the vector of student background characteristics (including baseline cognitive and socio-emotional assessment scores); δ_{cgs} and ϵ_{ict} ϵ_{icgs} is a residual. ANCOVA models, in which we control for baseline outcome measures, have more statistical power than post-treatment or double-difference estimators (McKenzie, 2012). To account for the fact that we have repeated post-treatment outcomes for the 2013 applicant cohort, we will cluster the standard errors when we estimate equation (1) at the applicant level. Estimates of ϕ capture the regression-adjusted ITT effect of being offered a first-grade slot at CHM. As a robustness check to estimates from regression equation (1), we propose to compute ITT effects using permutation-based methods, such as those proposed by Heckman, Moon, Pinto, Savelyev and Yalvitz (2010), Anderson and Legendre (1999), and Freedman and Lane (1983). Permutation-based methods are useful when sample sizes are small, creating covariate imbalance between treatment and control groups and when sample statistics are not normal. Heckman et al. (2010) and Anderson (2008), for example, have used permutation-based tests to analyze the impact of the HighScope Perry Preschool Program, an early childhood intervention for underprivileged minorities in Michigan, United States.

Following Angrist et al. (2010) we can also estimate the impact of CHM on outcomes as a function of time spent in CHM, which we can model as:

$$(2) \qquad Y_{ict} = \delta_c + \theta_t + \pi\, s_{ict} + X_{ic}'\rho + \varepsilon_{ict}$$

where δ_c and θ_t are cohort and year-of-measurement fixed effects; s_{ict} is the number of years applicant i from cohort c has spent in CHM at the time of the post-treatment measurement year t;[3] X_{ic} is the vector of student background characteristics (including baseline cognitive and socio-emotional assessment scores); δ_{cgs} and ϵ_{ict} ϵ_{icgs} is a residual. We can estimate equation (2) using an instrumental variables approach in which the randomly assigned lottery outcome Z_{ic} serves as an instrument for s_{ict}. As before, we cluster standard errors at the applicant level to account for the fact that we have repeated post-lottery outcome measures for the 2013 applicant cohort.

We will analyze program effects on multiple outcomes for children and parents. Some of these outcome domains include child cognitive development, early grade achievement, socio-emotional and parental outcomes, among others. Because testing multiple hypotheses increases the false discovery rate, we propose to follow the United States Institute of Educational Sciences' What Works Clearinghouse's convention in adjusting the hypotheses tests' critical p-values using the Benjamini-Hochberg (1995) adjustment.

[3] After we collect data in 2015, we have lottery losers from both the 2013 and 2014 cohorts likely spending zero academic years in CHM, lottery winners from the 2014 cohort spending up to one academic year in CHM and lottery winners from the 2013 cohort likely spending up to two years in CHM.

Randomization will ensure that, at baseline, selection bias will not be a threat to the internal validity of impact estimates. However, differential attrition between treatment and control groups can generate bias for the treatment effect estimates if those who attrit from the treatment group differ from those who attrit from the control group along dimensions related to the outcome of interest. In such a case, estimated program impact can be biased even if the lottery successfully generates comparable treatment and control groups at baseline.

The most important measure that we will take to reduce the risk of differential attrition between treatment and control groups will be to prioritize sample retention and high survey response rates in our follow-up data collection efforts, especially among control group applicants. We are confident that we will be able to retain the majority of baseline lottery applicants in both cohorts for a number of reasons. First, we will collect at baseline detailed parental contact information including address and phone numbers, as well as detailed contact information of close friends or relatives who can help us reconnect with original sample members. We will track lottery losers in the counterfactual schools and at home if necessary and aim to interview parents of lottery losers at home. Similarly, we will collect data from lottery winners at CHM and at home if necessary and, for parents of lottery winners, although we will aim to interview the majority of them at CHM's monthly parental meetings, we also plan to reach them at home, if necessary.

Second, we will offer incentives for survey response to parents of lottery losers.[4] Third, unlike other low-income families in smaller cities and rural areas, low-income families residing in Mexico City are not seasonal migrants and thus are not very geographically mobile. The average dropout rate for public elementary schools in Mexico City is 5.3 percent, lower than the 8.3 percent national average. Fourth, even if families of lottery losers move their child to another public school in Mexico City or other Mexican states, we can track them using administrative data from the Secretary of Education by linking these administrative records to lottery applicant data collected at intake that includes students' national identification numbers. However, because it is hard to ensure ex-ante that there will be no sample attrition, our power calculations also present a scenario with a 10 percent sample attrition rate.

Although we anticipate that with our data collection protocols sample attrition will be minimized, in the unlikely scenario that we detect it, we propose to approach it ex-post using cutting-edge statistical methods. Specifically, current "best practices" in the analysis of field experiments (Duflo, Glennerster, & Kremer, 2006) call for using two tests to determine whether attrition is likely to generate bias in program impact estimates. The first examines whether the fraction of applicants with missing outcome data differs between lottery winners and losers. The procedure closely resembles the ITT analysis of regression equation (1). Specifically, for each applicant cohort, we propose to regress A_{ic},

[4] Our budget assumes 200 Mexican pesos (around $US 15) gift cards for parents of control students to as a token of appreciation for completing follow-up surveys and tests. Gift cards will be obtained from popular grocery stores or bookstores. The minimum daily wage in Mexico City in 2012 was around $US 5. Therefore, these incentives are economically significant.

an indicator random variable for whether outcome data are missing for a given applicant, on assignment status and other previously defined covariates, as follows:

(3) $$A_{ic} = \vartheta_c + \eta_t + \tau\, Z_{ic} + X_{ic}{}'\vartheta + \zeta_{ic}$$

The null hypothesis of no difference in the probability of dropping out of the evaluation sample between lottery winners and losers corresponds to $H_0\colon \tau = 0$. Note that this analysis needs to be conducted for each outcome at each follow-up point (for example, one-year post lottery) since the attrition patterns can vary across outcomes and over time. The second test examines whether the factors that predict attrition are the same between lottery winners and lottery losers. To implement this test we modify equation (3) to include interactions between Z_{ic} and the vector of covariates X_{ic}:

(4) $$A_{ic} = \vartheta_c + \eta_t + \tau\, Z_{ic} + X_{ic}{}'\vartheta + Z_{ic}X_{ic}'\boldsymbol{\alpha} + \zeta_{ic}$$

and test the null joint hypothesis that all the parameters in the vector $\boldsymbol{\alpha}$ α are zero: $H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$, where m is the number of baseline characteristics in the vector X.

Rejection of the null hypothesis on either of these two tests raises concerns about attrition bias. If after data collection is complete we find evidence suggesting there may be bias from attrition, we propose to implement statistical procedures to estimate sharp bounds on the ITT effects that are valid even with attrition. Specifically, for continuous outcomes such as test scores we will employ the methods used in the Angrist, Bettinger and Kremer (2006) and Bettinger, Kremer and Saavedra (2010) evaluation of the effects of Colombia's PACES school voucher program. They obtained bounds on the ITT treatment effect by appropriately trimming the treatment group sample.

With regards to control-group contamination, we note that it can happen both directly, in the form of crossovers—lottery losers who gain future admission to CHM—and indirectly, in the form of treatment spillovers—lottery losers who benefit from lottery winners attending CHM. Spillovers from treated to control applicants are unlikely in this context. The most likely form of spillover effects would be among siblings. However, due to CHM's sibling priority policy, siblings of CHM students gain automatic admission to CHM. Therefore it is unlikely that lottery losers would benefit indirectly from the treatment of lottery winners. Control group crossovers are also unlikely to occur because admission to CHM—with one or two exceptions according to prior CHM administrative records—only takes place in first grade.

There are three main data sources for this evaluation. The first data source is baseline administrative data that CHM collects prior to the lottery from all lottery applicants, which includes student socio-demographic information, performance on the baseline screening tests that include, among others, the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV), and basic literacy tests.

The second data source is follow-up performance data on the Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA). These two tests will be our key outcomes of interest. EGRA measures students' ability to perform fundamental pre-reading and reading skills. It has been used to assess reading skills of children at grades 1-3 in more than 50 countries and 70 languages (Gove & Wetterberg, 2011). Prior studies showed it is a valid and reliable tool to measure reading achievement at early grade levels (RTI International, 2009). EGRA includes eight subtests, including letter name knowledge, phonemic awareness, letter sound knowledge, listening comprehension, unfamiliar non-word reading, familiar word reading, passage reading and comprehension, and dictation. Trained examiners administer each test to individual children orally. It takes about 15 minutes to complete the whole test. Results include eight subscale scores for analysis. Instruments, manuals, training videos and other information is available through the EDDATA II: Education Data for Decision Making website (https://www.eddataglobal.org/index.cfm). In addition the Mexican MoE has published National Standards for reading fluency to be used as a benchmark for this section of EGRA (http://www.consultasrodac.sep.gob.mx).

The third data source is cost data from CHM and public schools that lottery losers attend. To do so we will use a method to consistently identify value and distribute over time all costs, including costs of personnel, facilities, equipment and materials, other school inputs and client inputs. We propose to use the Ingredients Method of collecting cost data, which is a simple procedure to guarantee inclusion of all relevant costs (Dhaliwal, Duflo, Glennerster and Tulloch, 2011; Levin and McEwan, 2001, 2000). We will use an Ingredients Method excel worksheet template to collect comparable unit cost data across programs.

To appropriately analyze cost data we will document the assumptions required to model program costs at the margin (Dhaliwal, Duflo, Glennerster and Tulloch, 2011). Quantification of program costs requires a "comparator case"—a baseline cost scenario with which to compare treatment costs. The relevant cost comparator scenario depends on the context in which the program might be replicated or scaled. For example, for CHM, the cost comparator case will likely be a situation in which control students attend public schools in Mexico City, so that the cost at the margin is only the cost of CHM. To better understand the cost comparator case, we will use the same Ingredients Method to collect cost data from the two public schools that in our sample concentrate the most number of CHM lottery losers.

An important consideration when collecting and analyzing cost data is to account for all relevant costs. For example, CHM receives certain goods and services as donations. There are other costs that CHM accrues that have overlapping uses, such as use of instructional facilities (Dhaliwal, Duflo, Glennerster and Tulloch, 2011). There is also the additional cost of parental time due to mandatory school involvement activities at CHM. To accurately analyze cost data, costs associated with facilities should be adequately annualized (Levin and McEwan, 2000). Similarly, donations and overlapping uses should be appropriately accounted for to the extent that they are relevant costs to society if the programs were to be adopted at a larger scale. We will be very explicit about whether we include or exclude donating, overlapping and other potentially relevant costs—financial and in-kind—and will justify our decision for each cost category. Among project deliverables, we will provide all disaggregated cost data, organized with our excel template to facilitate re-estimating cost-effectiveness measures under alternative scenarios.

It is possible that CHM impacts participants in different ways. For example, CHM might be more beneficial for girls than for boys or for single-parent children. Accounting for outcome heterogeneity is crucial to rigorous cost-effectiveness comparison (Levin and McEwan, 2000). Therefore we will carry out all impact analyses (and corresponding evaluation designs) using sex-disaggregated data.

One important consideration when analyzing cost data concerns how to convert all impact and cost measures to common units, therefore for this evaluation's cost analysis, we propose to sequentially perform the following steps: (i) convert yearly cost data to U.S. dollars using year-specific exchange rates, (ii) deflate each program's costs to the corresponding base year using average annual United States inflation rates, (iii) obtain the present value of costs using a 10 percent discount rate, and (iv) convert to year-of-analysis United States dollars using annual United States inflation rates between the base year and the year of analysis (Dhaliwal, Duflo, Glennerster and Tulloch, 2011).

Data will be collected annually from FY 2013 to FY 2016. Data collection will be conducted annually from the baseline year (that is, the year before students are admitted into CHM) to the end of 2nd grade (see Table 1). We plan to collect data during May of each evaluation year.[5]

[5] The school calendar in Mexico runs from late August to early July.