MATHEMATICA
CENTER FOR
INTERNATIONAL POLICY
RESEARCH AND EVALUATION

Evaluation Design Report for the Georgia Improving General Education Quality Project's School Rehabilitation Activity

Final Report

March 5, 2015

Ira Nichols-Barrer
Caroline Lauver
Leigh Linden
Matt Sloan

MATHEMATICA
Policy Research

MATHEMATICA
CENTER FOR
INTERNATIONAL POLICY
RESEARCH AND EVALUATION

Contract Number:
MCC-13-BPA-0040 (CL-002)

Mathematica Reference Number:
40306.200

Submitted to:
Millennium Challenge Corporation
875 Fifteenth St., NW
Washington, DC, 20005
Project Officer: Ryan Moore

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Matt Sloan

Evaluation Design Report for the Georgia Improving General Education Quality Project's School Rehabilitation Activity

Final Report

March 5, 2015

Ira Nichols-Barrer
Caroline Lauver
Leigh Linden
Matt Sloan

MATHEMATICA
Policy Research

CONTENTS

TABLES

FIGURES

1. Introduction

The Millennium Challenge Corporation (MCC) is supporting Georgia's efforts to improve educational outcomes by sponsoring the Improving General Education Quality (IGEQ) Project, which includes three components. The Improved Learning Environment Infrastructure (ILEI) component invests in school rehabilitation to provide safe learning environments that include adequate facilities and heating. The Training Educators for Excellence (TEE) component supports professional development by training and mentoring teachers to improve competencies in science, technology, engineering, and math subjects and by training principals to strengthen school management. Finally, the Education Assessment Support (EAS) component supports Georgia's ongoing efforts to improve educational outcomes through rigorous assessments and fostering a result-oriented education system. Mathematica Policy Research is designing and implementing a rigorous evaluation of these components to determine their ultimate impact on both intermediate and long-term outcomes.

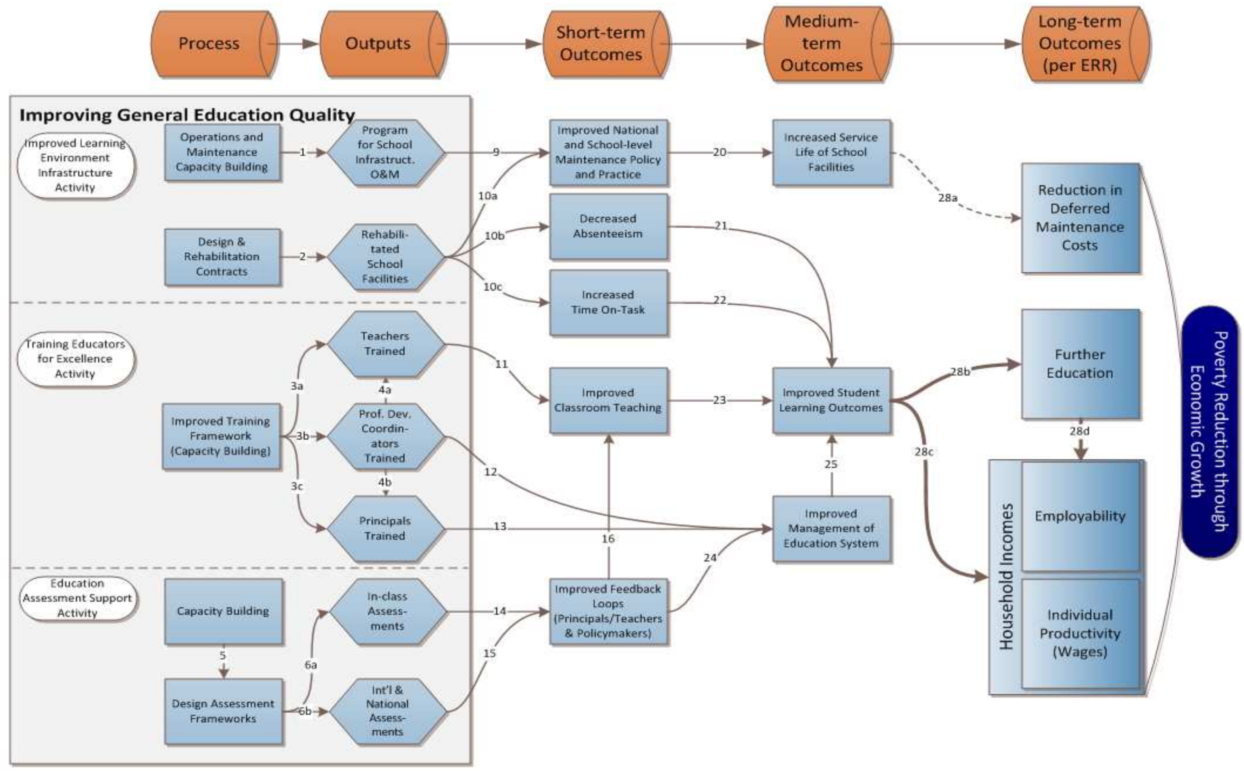
This design report provides a detailed explanation of the evaluation design chosen for the ILEI (school rehabilitation) activity. When program implementation plans have been developed for the TEE and EAS activities, we will prepare a new design report including the evaluation designs chosen for those two activities as well. We begin by presenting an overview of the program logic for each component of the IGEQ intervention and briefly review the existing literature examining the impacts of similar interventions in other countries. Next, we present a detailed explanation of the ILEI activity's evaluation design, providing a discussion of key evaluation questions, methods, and data sources for the study's major outcomes.

2. Overview of the school rehabilitation activity design

The school rehabilitation activity is designed to upgrade the quality of physical infrastructure and create an improved learning environment in program schools. Examples of potential rehabilitation areas include systems for heating (replacing wood stoves with central heating); lighting; water and plumbing; lavatories; recreational facilities; science laboratories; building interiors (flooring, stairs, and classroom walls); and building exteriors (roofing and masonry). The activity plans to rehabilitate up to 130 schools throughout Georgia and the work is scheduled to take place over the course of three construction seasons (the summers of 2015, 2016, and 2017).

According to the program's logic model (Figure 2.1), these inputs are intended to decrease students' and teachers' absenteeism and improve time on task during the school day, leading to improved student learning and higher educational attainment outcomes. Although it is not reflected in the program's current logic model, we also believe it is plausible that rehabilitating schools could improve the health and well-being of students, which might provide another pathway for the intervention to affect learning and other long-term outcomes. The program logic developed by MCC and Millennium Challenge Account-Georgia (MCA-G) staff presents a series of (hypothesized) causal links among program inputs and outputs and short-, medium-, and long-term outcomes that potentially support the project's overarching goal of poverty reduction through economic growth. Each of the links in the program logic represents an assumption by IGEQ program designers about how the activities will affect the compact's beneficiaries and stakeholders, which include students, teachers, school administrators, and policymakers in relevant Government of Georgia (GoG) ministries and centers. Assumptions in the program logic also provide the basis for MCC's economic rate of return (ERR) calculations for each activity.

Figure 2.1. The IGEQ program logic



Source: MCC Georgia II Compact Investment Memo.

Note: Arrows with dotted lines refer to links that MCC does not expect to be evaluable or measurable. "O&M" refers to operations and maintenance expenses.

To assess the IGEQ program logic and associated ERR calculations, we reviewed the available evidence on the impacts of similar program designs in other contexts and held detailed discussions with local education experts and IGEQ stakeholders during the Mathematica team's initial trip to Georgia in November 2013. These discussions included MCA-G staff, stakeholders in relevant GoG centers and ministries, and the team's site visits to schools selected for the ILEI rehabilitation program. We examined the program logic for each of the three components of the IGEQ separately, noting potential concerns where applicable in a logic assessment report (Nichols-Barrer et al. 2013). Our review of the relevant literature is summarized in the next section.

3. Literature review

There is an extensive academic literature investigating the relationship between educational inputs and measures of student learning, educational attainment, and employment outcomes. However, much less is known about the impacts of education interventions in developing countries, and little empirical work exists on the education system in Georgia.

Based on the Mathematica research team's initial evidence review and discussions with program stakeholders, in general, the program logic for the ILEI activity represents a plausible set of assumptions regarding how improved school infrastructure could lead to improved student outcomes and a reduction in schools' deferred maintenance costs. However, the existing evidence base does not support strong predictions about the size of a program's expected impacts on some of the key outcomes. We provide an overview of the relevant literature below.

According to the ERR calculations used for the school rehabilitation activity, MCC aims for this intervention to produce the following improvements in students' long-term outcomes: a 10 percent improvement in the number of students enrolling in upper secondary school; and a 10 percent improvement in postsecondary enrollment rates. In terms of evidence from prior studies, there is great uncertainty regarding the relationship between school infrastructure inputs and all of these aforementioned outcomes. Some evaluations of school construction and rehabilitation activities have found positive impacts on students' enrollment and attainment in some contexts (Burde and Linden 2013; Levy et al. 2009; Durán-Narucki 2008; Woolner et al. 2007) and limited to no impact in other contexts (Dumitrescu et al. 2011). There is very limited rigorous research that assesses whether there is a causal link between school rehabilitation inputs and long-run improvements in employment rates or income levels; in fact, we are not aware of any studies that have tested this question using reliable empirical methods in developing countries. Measuring these long-term outcomes as part of an extended evaluation study would be a substantial contribution to the research literature and fill a significant gap in knowledge.

A major focus of past studies on school infrastructure has been the relationship between school-building interventions or infrastructure improvements and student attendance. Specifically, researchers have tested whether attendance rates improve following upgrades to school infrastructure. Several studies in both domestic and developing country contexts have shown that improving schools' physical infrastructure can lead to an increase in school enrollment and attendance. However, the impacts of infrastructure improvements likely depend on preexisting conditions in the affected facilities or communities. For example, if a program improves a school that is already functioning well, one would expect the benefits of the program to be relatively modest. Conversely, in a community with very limited school facilities, construction or rehabilitation programs can produce large benefits.

For example, an impact evaluation of the BRIGHT I program in Burkina Faso, an initiative that constructed primary schools in 132 rural villages throughout the 10 provinces with the lowest girls' school enrollment rates, specifically targeted communities that did not previously have ready access to a school. The evaluation found that BRIGHT I schools had a positive impact on school enrollment and a large impact on math and French test scores for both boys and girls (Levy et al. 2009). Several descriptive studies of school conditions in the United States have found analogous results. A study in New York City examining the relationship between poor school facilities and various student outcomes found that students in the most deteriorated buildings attended fewer days of school and had lower test scores in English language arts and mathematics (Durán-Narucki 2008). A pre-post case study on the effects of the renovation of a run-down elementary school in Washington, DC found evidence of improved student attendance and test scores (Berry 2002). However, other studies show that investment in schools' physical infrastructure does not necessarily improve student attendance. The IMAGINE program in Niger constructed schools in 10 communities with low enrollment and primary school completion rates for girls, but—unlike the BRIGHT I program implemented in Burkina Faso—many of these areas already had an existing school. Although the study did find that the newly constructed schools raised enrollment by 4.3 percentage points, there was no short-term impact on attendance rates, math test scores, or French test scores (Dumitrescu et al. 2011). A forthcoming follow-up study will also examine the program's longer-term impacts.

Substantial evidence suggests that increases in the amount of time students spend on learning tasks in school can improve their test scores. However, few studies have examined the impacts of infrastructure on time use during the school day, and it is not clear whether school building improvements lead to increases in the hours of functional instruction students receive. That said, if we assume (as shown in the rehabilitation activity's logic model) that the intervention could increase learning time, evidence suggests that, in turn, this could produce important learning gains.

Studies in the United States and developing countries provide evidence that additional time spent on learning tasks can plausibly improve students' test scores. For example, a randomized evaluation on the effects of short-term tutoring on cognitive and non-cognitive skills in Chile found that students from low-performing and poor schools improved their reading test scores after participating in the three-month program (Cabezas et al. 2011). Similarly, a participatory program in India trained local village volunteers on pedagogical techniques for teaching basic reading skills and subsequently tasked them to hold daily reading classes outside of school in an effort to improve the learning of village children. A randomized evaluation of the program found the additional instruction had a positive effect on the reading skills of children who attended the camp (Banerjee et al. 2010). A great deal of research in the United States has also examined the relationship between the amount of instructional time and student learning. Studies of New York City charter schools have found that high-achieving charter schools tend to have a longer instructional year and longer school days than other charter schools (Hoxby et al. 2009; Dobbie and Fryer 2013). One of these studies found that these characteristics, coupled with frequent teacher feedback, data-driven instruction, and a focus on academic achievement, explained almost half of the variation in school effectiveness (Dobbie and Fryer 2013). A national study of the relationships between the practices of individual charter-school management organizations (CMOs) and their effects on student achievement found that CMOs with lengthened instructional hours (alongside school-wide behavior policies and more intensive teacher coaching) had larger impacts on student achievement in math and reading than other categories of CMOs (Furgeson et al. 2012).

It is important to note that none of the strong prior studies on school infrastructure specifically address the context in Georgia. Without evidence and knowledge on the determinants of enrollment, attendance, achievement, and attainment in the Georgian context, it is difficult to predict whether infrastructure improvements in Georgian schools will have a positive effect on student outcomes. Likewise, although studies in other countries suggest that increased time on task can have a positive effect on student learning, it is unclear whether in the Georgian context teachers will be able to use additional instruction time effectively to raise student test scores.

4. Evaluation design for the ILEI activity

The school rehabilitation activity seeks to decrease student and teacher absenteeism, increase students' time on task, and, ultimately, improve learning and labor market outcomes. This section describes our evaluation design for assessing how the ILEI activity is implemented and estimating its impacts on these outcomes.

4.1. Evaluation type

We propose a mixed-methods study design, with three components: (1) a process evaluation examining the program's implementation and costs; (2) a randomized controlled trial (RCT) impact evaluation using a school-level stratified random assignment design, and (3) in-depth analysis of the relationship between changes in school infrastructure and changes in the learning environment, using qualitative methods in a subset of study schools.

4.2. Evaluation questions

Table 4.1 presents the key research questions to be investigated. Our process evaluation will examine outcomes related to program design and implementation, and the impact evaluation and in-depth qualitative analyses will examine the program's effects on school infrastructure, teachers, and students. The table also summarizes the data sources we will use for each research.

Table 4.1. Evaluation questions for the ILEI activity and approaches to answering them

Key evaluation questions	Evaluation components
Program design and implementation	**Process evaluation**
Was the ILEI activity budgeted and planned appropriately, forecasting key risks?	• Compare implementer's projected and actual cost data and examine risk assessment documents
Did the ILEI activity deliver improved facilities? How was the program rolled out? How much did rehabilitation differ by school?	• Use implementer data to compare time lines, budgets, work plans, and material use
What is the current and future status of facility-maintenance funding for schools? Do treatment schools have ongoing operations and maintenance funding to use in improved facilities? What maintenance/rehabilitation funding did control schools receive?	• Interview school directors to gather data on operations and maintenance funding and maintenance practices • Review GoG budget allocation methods to schools as they pertain to operations costs
Impacts on infrastructure, teachers, and students	**Impact evaluation (RCT) and qualitative analysis**
What are the impacts of the ILEI activity on the school infrastructure environment, such as temperature, maintenance policy, and maintenance practice? Did the activity affect perceptions of student and teacher health and safety?	• Assess quality of school facilities, including observational data from enumerators on temperatures during the school day; conduct surveys and in-depth interviews with school directors regarding operations practices and equipment usage
What are the impacts of the ILEI activity on teacher behavior, such as attendance and time spent teaching?	• Analyze teacher and student survey data; conduct in-depth interviews with teachers and student focus groups
What were the impacts of the ILEI activity on student outcomes? What are the impacts on attendance, enrollment, drop-out and retention rates, time spent studying in and out of school, and learning outcomes?	• Analyze teacher and student attendance through school visits (preferred) or administrative data; analyze time on task and teaching practices through classroom observation (video) data • Analyze student test scores
Impacts on attainment and employment	**Impact evaluation (RCT)**
What are the long-term impacts of the ILEI activity? What are the impacts on school-level student attainment (transition to secondary school and secondary school graduation) and on teacher qualifications at rehabilitated schools?	• Analyze administrative data on student attainment rates and teacher qualifications • If the study is extended beyond 2019, a long-term follow-up survey of students could examine postsecondary attainment and employment outcomes

4.3. Methodology

In this section, we explain the methods associated with each component of our evaluation for the school rehabilitation activity.

Process evaluation examining program implementation and costs

For the process evaluation, Mathematica will begin by reviewing ILEI activity documents, including program cost data, program implementation records, and school rehabilitation design assessment reports, as available. These reports should document site assessments, rehabilitation recommendations, and implementation records for the program's treatment schools. From this, we can develop a basic understanding of program implementation and inputs.

We will supplement the document review by conducting a series of in-depth, semi-structured interviews targeting three groups of respondents: key GoG staff, implementers including the activity's design contractor(s), and rehabilitation supervisors. We will develop the interview guides around numerous themes that will include, but not be limited to, respondent knowledge, attitudes, perceptions, and commitment to the ILEI activity; documentation and impressions of implementation activities; specific barriers to and challenges with rehabilitating schools; and suggestions on alternative strategies for supporting school rehabilitation efforts. We will use the major topics and themes that emerge from the review of program documents to help develop these semi-structured interview protocols. We will use these data to examine implementation successes and challenges and to document key lessons learned about implementation of school rehabilitation programs, as well as implications that could help inform implementation of similar programs in other contexts.

Impact evaluation applying an RCT design

To estimate the impacts of the school rehabilitation activity, our study uses a school-level, stratified random assignment design. Schools assigned to the treatment group will at minimum receive detailed rehabilitation design assessments, and—where rehabilitation is feasible—treatment schools will receive the program's full set of infrastructure rehabilitation services. As part of the Georgia II Compact agreement, GoG stakeholders have agreed that schools assigned to the control group will only receive "business as usual" maintenance and operations support during the life of the five-year compact (until July 2019).

To develop the random assignment procedure, the design first stratifies the sample of schools by region. Within regions, we then consider the benefits of further stratifying the sample on the following school-level characteristics:

- Total enrollment
- Secondary enrollment (students in grades 10 to 12)
- Size of school building
- Government rating of school infrastructure conditions
- Minority language status (indicator for instruction primarily in Azeri or Armenian)
- Rural status (indicator for school located in a village or mountainous area)
- Average baseline test scores in math, history, and literacy

In addition, the stratification approach takes into account the design status of schools in the sample. During the 2013-2014 school year (before random assignment), MCA-G hired a design contractor (Louis Berger) and partially or fully completed rehabilitation designs for several schools in program regions. Due to implementation delays, no rehabilitation work took place in these schools during the 2014 summer construction season, meaning the predesigned cases could be included in the random assignment pool for this evaluation. In total, 29 program-eligible schools have existing rehabilitation designs. To realize cost savings from this prior design work, at the request of MCA-G and MCC, the evaluation will give the predesigned schools a higher probability of being assigned to treatment (66 percent) than the schools currently lacking designs. To do so, our approach places the pool of predesigned schools in its own separate set of random assignment blocks. The study's impact analyses will adjust statistically for differences in the probability of selection into treatment associated with these predesigned strata.

This random assignment process will take place in three phases corresponding to the program's three years of implementation. Each of Georgia's regions has been assigned to a different phase (Table 4.2)—this enables the rehabilitation work in each phase to take place in a set of regions that are close to each other, facilitating program logistics. Due to this staggered rollout schedule, the evaluation will acquire complete baseline data on the full evaluation sample (that is, covering schools in all three phases) during the 2016-2017 school year, and the study will complete its first full analyses of the program's Year 1 follow-up impacts after data is collected during the 2017–2018 school year.

Table 4.2. Regional rollout of the ILEI activity

Phase	Regions	Approximate number of treatment schools	Schedule for rehabilitation
I	Mtskheta-Mtianeti, Racha-Lechkhumi and Kvemo Svaneti, Samtskhe-Javakheti, Shida Kartli	30	Summer 2015
II	Kakheti, Kvemo Kartli	35	Summer 2016
III	Adjara, Guria, Imereti, Samegrelo-Zemo Svaneti	40	Summer 2017

At the beginning of a given phase, Mathematica will randomly select which schools are eligible to receive the program from a list of suitable schools in each region that has been vetted by MCC, MCA-G, and GoG stakeholders. Mathematica completed the random assignment process for schools in the Phase I regions in September 2014.

In-depth qualitative research on the effects of school rehabilitation

In addition to the process evaluation, our approach also includes qualitative research designed to enrich the study's quantitative impact analyses by generating hypotheses about how school rehabilitation changes the learning environment and student outcomes. Qualitative methods provide a means of investigating potential mechanisms responsible for driving the program's impacts by collecting the type of extensive, open-ended interview and focus group data that would not be feasible to collect and analyze in all study schools. The proposed qualitative analysis will be implemented by Mathematica in conjunction with a local data collection firm. We propose to collect qualitative data in the second follow-up year after rehabilitation in each treatment school. In total, Mathematica will select a subset of approximately 10 percent of the schools in the impact evaluation sample (20 schools—10 treatment and 10 control), and the local data collection firm will collect in-depth, qualitative data about program implementation and results at these schools. The data collection will pay particular attention to maintenance and operations practices, perceptions of school quality and safety, time on task, and the use of various school facilities. This information will be acquired by conducting in-depth interviews with school directors and teachers and by conducting focus group discussions with secondary school students. The in-depth interviews with school directors will assess infrastructure usage patterns, school operations, and maintenance practices; the in-depth interviews with teachers will assess how school facilities are used, time on task, perceptions of school building quality and safety, and teacher attendance; and the focus groups with students will likewise assess how school facilities are used, time on task, perceptions of school quality and safety, and determinants of student attendance.

We expect insights from these qualitative research activities to be important and valuable, but it is important to note that qualitative methods have certain limitations. As with most qualitative research, findings from stakeholder interviews and focus groups will be illustrative and do not have the sample size to support rigorous hypothesis tests to directly estimate the program's impacts on the population being studied. We will focus on capturing how the activity was implemented, gaining an understanding of a broad set of implementation issues from a diverse set of stakeholders, and investigating the ways that school rehabilitation might affect teachers and students to improve attendance and learning outcomes. From these data, it will be possible to draw some conclusions about the potential reasons for the pattern of impacts uncovered by the impact evaluation, lessons learned in relation to implementation strategies and their potential to support school rehabilitation projects, and the potential relationships between various school infrastructure inputs and key program outcomes.

4.4. Study population

The evaluation will focus on estimating the impacts of school rehabilitation on students and teachers. In particular, the evaluation's findings will pertain to the population of students enrolled or potentially enrolled in the types of schools selected to receive rehabilitation services. The evaluation's primary sample of interest will be the population of students enrolled in grade 8 or grade 10 at baseline in study schools—we will track these two cohorts of students over time and measure their outcomes during two follow-up years. The characteristics of the sample of eligible schools are summarized in Table 4.3.

To identify the program sample, MCC and MCA-G developed recommendations regarding the percentage of program schools that should be allocated to each of Georgia's regions, excluding schools in major urban areas such as Tbilisi and Batumi. Within each region, schools were then ranked on a set of eligibility criteria including having high enrollment, high space utilization rates, poor building conditions, and a high percentage of enrolled students who are classified by the government as socially vulnerable. However, to maximize the cost-effectiveness of the program (that is, the number of benefiting students per dollar spent), MCC requested that the ranking approach place a greater weight on space utilization than other criteria. As shown in Table 4.3, this ultimately produced a study sample with higher average enrollment, higher space utilization rates, and lower percentage of rural schools and socially vulnerable students than the national average in Georgia.

These criteria were used to identify an initial list of 425 potential schools. Next, MCA-G and GoG stakeholders reviewed this list and recommended that 108 schools be removed from consideration due to issues such as major structural faults (which are not cost effective to address), environmental risks, or concerns over unclear land titles. Of the remaining list of 317 eligible schools, schools with the highest utilization rates within each region were selected to form a sample of 184 schools eligible for random assignment, with the number of schools in each region being allocated in the same proportion as previously agreed (this initial list did not include Adjara, but eventually up to 14 schools may also be deemed eligible in the Adjara region when eligibility criteria for that region have been established). The schools with lower utilization rates in each region were used to form a separate list of 117 reserve-pool schools that may be added to the program sample for random assignment at a later date due either to programmatic cost savings (allowing more schools to be rehabilitated) or exclusion of treatment schools following detailed design assessments.

Summary characteristics for the rehabilitation program's initial pool of 184 eligible schools are shown in Table 4.3.

Table 4.3. Characteristics of schools eligible for random assignment to rehabilitation services

	Evaluation sample	Georgia
Number of schools	184	1,692
Average total enrollment	414.8	173.4
Average secondary level enrollment (grades 7–12)	197.8	86.2
Ratio of school building size (m^2) to student enrollment	6.2	16.7
Government rating of building condition (0–10 scale)	5.0	5.5
Percentage of socially vulnerable students	20	27
Percentage of students in Azeri language schools	8	8
Percentage of students in Armenian language schools	3	5
Percentage of students in mountain or village schools	40	64
Regional distribution of schools (percentage)		
Adjara (eligible schools to be determined)	8	12
Guria	4	5
Imereti	20	20
Kakheti	20	10
Kvemo Kartli	16	14
Mtskheta-Mtianeti	2	4
Racha-Lechkhumi and Kvemo Svaneti	3	3
Samegrelo-Zemo Svaneti	8	12
Samtskhe-Javakheti	6	11
Shida Kartli	14	8

Note: The sample of schools summarized in this table excludes schools in the cities of Tbilisi and Batumi (because urban areas are not eligible for the program) and schools outside Batumi in the Adjara region (because eligibility criteria in that region are still being negotiated).

4.5. Study sample and power calculations

To align the data collection with the key outcomes envisaged in the ILEI activity's program logic, we propose to target data collection efforts to students in both lower- and upper-secondary grades. Specifically, in each school we will define the study sample as all students enrolled in grades 8 and 10 in the school's baseline year. By tracking these two student cohorts longitudinally (regardless of whether they remain enrolled in the same school, transfer to a different school, or drop out of school altogether), we will be able to observe the impacts of rehabilitation activities on key secondary-school enrollment and attainment outcomes. For example, we will be able to observe students' transition rates from lower-secondary to upper-secondary grades (because students in grade 8 at baseline will have had time to reach the beginning of upper-secondary school in grade 10 by the second follow-up year). Likewise, students enrolled in grade 10 at baseline will have reached grade 12 by the second follow-up year, potentially enabling us to observe the program's impacts on national secondary school exit examinations and secondary school completion. Under this approach, we estimate that the sample will include approximately 60 students in each school.

We present power calculations for the study in Table 4.4, showing the statistical precision provided by four illustrative sample configurations. In the benchmark scenario, we calculate the power of the study assuming treatment and control groups of 100 schools each. Note, however, that the final number of treatment and comparison schools in the sample has yet to be determined (for example, some schools originally selected for the program might be excluded for eligibility reasons or the number of program schools could decline if there are unexpectedly high rehabilitation costs). For the purpose of defining these scenarios, we assume that each non-rehabilitated school in the treatment group will be replaced by expanding the evaluation sample to include an additional treatment and control school (for a more detailed discussion of these exclusion and replacement scenarios, see Section 7 of this report). To reflect these possibilities, the power calculations show a variety of scenarios regarding the number of treatment and comparison schools that will ultimately be included in the evaluation sample. Specifically, we show power calculations for different scenarios reflecting the rate at which schools initially assigned to the treatment group could be classified as ineligible for the program.

Depending on the final number of schools that are included in the ILEI activity, we estimate that the evaluation will be able to detect statistically significant student-level impacts as small as 0.11 standard deviations in the best case and 0.16 standard deviations in the least favorable case.

Table 4.4. ILEI minimum detectable effects (MDE) for different sample sizes and compliance rates

	All schools in the treatment group are rehabilitated		Some treatment schools are not rehabilitated (each initial exclusion replaced with one treatment and one control school)	
	Benchmark program size	Reduced program size	20 percent of treatment schools excluded	40 percent of treatment schools excluded
Evaluation sample of schools	100 treatment 100 control	80 treatment 80 control	120 treatment 120 control	140 treatment 140 control
Number of rehabilitated schools	100	80	96	84
Total student sample	12,000	9,600	14,400	16,800
MDE for all schools assigned to treatment (ITT impacts)	0.11	0.13	0.10	0.09
Compliance with treatment group assignment	100%	100%	80%	60%
MDE for rehabilitated schools (TOT impacts)	0.11	0.13	0.13	0.16

Notes: Sample sizes assume that for each of the initially assigned treatment schools that is not rehabilitated, the evaluation sample will expand by adding one treatment school and one control school. MDE calculations assume a two-tailed test with a 5 percent significance level and 80 percent power. We assume an intraclass correlation (ICC) of 0.1, a school-level R-squared of 0.3, a student-level R-squared of 0.2, and an aggregate student sample comprising 30 students in grade 8 and 30 students in grade 10 enrolled at baseline in each study school. MDE calculations for cohort-specific outcomes assume a reduced sample size of 30 students per school. The ICC and R-squared assumptions are based on U.S. data from school-level cluster randomized trials in education, as reported in Hedges and Hedberg (2007) and Deke et al. (2010). Treatment-on-the-treated (TOT) MDEs were calculated by dividing the intent-to-treat (ITT) MDEs by the compliance rate among treatment schools (this assumes no control schools receive treatment).

Based on our review of other school construction evaluations in developing countries, we believe that the range of detectable effects shown in these scenarios represents a level of statistical precision that is adequate to detect impacts comparable to those reported for school construction in certain other contexts (Levy et al. 2009). However, it is important to note that school construction interventions have not always produced sizeable short-term impacts (e.g., Dumitrescu et al. 2011), and that prior studies have tended to examine wholesale construction of new school buildings rather than rehabilitation of existing facilities. Even with a minimum detectable effect equal to 0.11 standard deviations (the best-case scenario shown in Table 4.4), we cannot say with confidence whether the evaluation will find significant impacts.

For the process evaluation of school rehabilitation activities, we recommend conducting a series of in-depth interviews targeting three groups of respondents: 1 or 2 key GoG staff, 1 interview with each of the activity's design contractors, 2 interviews with rehabilitation supervisors, and 2 interviews of MCC/MCA staff involved in implementation and oversight of the rehabilitation program. We believe that collecting information from the respondents involved in each area of activity implementation will enable us to develop a full picture of the planned implementation, the actual implementation, and the reasons for any divergences between the planned and actual implementations.

At a subset of treatment schools, we also recommend an in-depth data collection for an analysis that will use descriptive and qualitative methods to investigate how rehabilitation affected the learning environment at study schools. We recommend drawing a sample designed to obtain representative information from each of the program's 10 geographic regions in the study's second follow-up year., We propose to include a sample of two schools in each region—one treatment school and one control school—in this additional data collection effort. Across regions, schools will be purposively selected to include a representative range of characteristics, such as school size and urbanicity. Within each of these schools, the local data collection firm will conduct one in-depth interview with the school director, in-depth interviews with four teachers (including at least one science teacher), and two student focus groups. Each focus group will include approximately eight randomly selected students in secondary-level grades. We will consider stratifying the student focus groups either by gender or by grade level (upper secondary grades versus lower secondary grades) when the qualitative data collection is piloted in 2017. Survey data from the 2015 and 2016 data collection rounds may also shed light on whether there are gender or grade-level differences with respect to school rehabilitation that merit further investigation through focus groups. In total, the qualitative sample will consist of 20 schools providing a total of 20 school director interviews, 80 teacher interviews, and 40 student focus groups. Although we believe these samples will produce meaningful descriptive data for qualitative analysis, this subsample of schools is too small to support quantitative hypothesis testing, and, as a result, we do not show power calculations for this portion of the study.

4.6. Time frame

Each of the quantitative data collection components will be collected from all treatment and comparison schools at three points in time: the baseline year when random assignment occurs, the first follow-up year after rehabilitation is assigned to begin, and the second follow-up year after rehabilitation is assigned to begin. Tracking outcomes for two follow-up years will enable us to examine immediate impacts shortly after rehabilitation work occurs and two-year impacts examining the program's longer-run effects. The timing of the follow-up data collection rounds is designed to coincide with important academic transitions for the student cohorts that will be tracked longitudinally in each treatment and control school. After two years, the grade 8 cohort will have reached the beginning of upper secondary school, and by then the grade 10 cohort will have reached the end of upper secondary school (grade 12). Thus, conducting a one-year follow-up data collection will measure the program's immediate impacts, and the two-year follow-up will examine whether these effects persist after students transition into upper secondary school or complete their secondary education. We also propose the local data collection firm conduct qualitative data collection activities in a subsample of schools in the second follow-up year after rehabilitation begins.

The data collection contract will be structured with an initial one-year period of performance covering baseline data collection in 2015, followed by four additional options for follow-up data collection rounds in 2016, 2017, 2018, and 2019 (Table 4.5). Because ILEI rehabilitation activities will occur in three phases, the data collection rounds will occur in the following sequence (data collection for a given phase encompasses all treatment and comparison schools in the regions assigned to that phase). Note also that in 2018, following the end of construction in Phase III schools, Mathematica will collect additional process evaluation data beyond the surveys, student learning assessments, and qualitative data collected across the other data collection rounds. For example, for the process evaluation the study would collect all available ILEI implementation reports and cost records after completion of rehabilitation work in Phase III.

Table 4.5. Data collection schedule

Collection round (March-June of each school year).	**Phase I regions** (Mtskheta-Mtianeti, Racha-Lechkhumi and Kvemo Svaneti, Samtskhe-Javakheti, Shida Kartli)	**Phase II regions** (Kakheti, Kvemo Kartli)	**Phase III regions** (Adjara, Guria, Imereti, Samegrelo-Zemo Svaneti)
2015	Baseline data collection with grade 8 and 10 students	None	None
2016	One-year follow-up with grade 9 and 11 students	Baseline data collection with grade 8 and 10 students	None
2017	Two-year follow-up with grade 10 and 12 students. Qualitative data collection	One-year follow-up with grade 9 and 11 students	Baseline data collection with grade 8 and 10 students
2018	None	Two-year follow-up with grade 10 and 12 students. Qualitative data collection	One-year follow-up with grade 9 and 11 students
2019	None	None	Two-year follow-up with grade 10 and 12 students. Qualitative data collection

Due to this staggered approach to data collection, the 2017 data collection round will be the only year in which the local data collection firm conducts site visits in *all* of the evaluation's treatment and comparison schools. Thus, in most years the number of visited schools will be smaller, potentially reducing the logistical burdens associated with the data collection effort.

If the ILEI implementation plan changes, the study team will consider appropriate revisions to the data collection schedule. For example, if the program elects to add a fourth construction season in summer 2018, the evaluation team would lead discussions with MCC and other stakeholders to assess how best to measure outcomes in the schools that are rehabilitated in 2018 (e.g., the evaluation could consider measuring one-year follow-up outcomes in these schools in the 2019 data collection round, and stakeholders could consider pursuing an additional data collection round for these schools in 2020).

Likewise, we will consider alternative or extended data collection schedules as the program develops. In coordination with MCA-G, the study team has recommended the use of a year-by-year contract with the local survey firm and the use of a year-by-year implementing entity agreement with NAEC covering student learning assessments. This approach will provide an opportunity to assess whether the existing data collection plan is still advisable following each data collection round, since the contract structure facilitates making adjustments on a yearly basis. For example, after the 2016 round (which includes the first-year follow-up in Phase I schools) we could consider adjusting the timing of the second follow-up data collection round or adding additional data collection rounds to the study. The study team will maintain a flexible approach, and will discuss the merits of changes to the study design and data collection plan with MCC, MCA-G and other stakeholders as needed throughout the life of the study.

5. Data sources and outcome definitions

Our design calls for collection of survey data on the ILEI activity's key outcomes from students, parents, teachers and school directors. This will be complemented by a combination of administrative data, study-administered learning assessments, direct observations of student attendance and school infrastructure, and qualitative research. Survey data, learning assessments, direct observations of attendance, and ratings of school infrastructure will be collected by a MCA-procured local data collection firm. Mathematica will obtain administrative data from Georgia's education management information system (EMIS) and activity implementation records.

The data sources for each of the study's key outcomes are summarized in Table 5.1.

Table 5.1. Data sources and study outcomes for the ILEI evaluation

Component	Description	Outcome
Data to be collected directly by Mathematica		
Administrative EMIS data	Longitudinal, student-level records detailing school of enrollment and grade of enrollment for the study's student sample in all study years. Administrative data might also provide information on student and/or parent characteristics. If possible, we would seek to merge EMIS data with national assessment data to obtain students' test scores on the national secondary-school exit examination.	National assessment test scores in math, science, and literacy Student transition rates from lower- to upper-secondary grades (10, 11, and 12) Student drop-out/retention rates Secondary school completion rates Transfer rates to/from nearby schools Experience of teaching staff; percentage of certified teachers Total school enrollment (all grades), total enrollment in elementary school, total enrollment in lower secondary school, and total enrollment in upper secondary school.
Rehabilitation design and implementation records	To document the design and early implementation of the ILEI activity, Mathematica will obtain any available school rehabilitation design assessment reports, additional program implementation records, and program cost data.	Process analysis

Component	Description	Outcome
In-depth interviews with implementers	For the study's process evaluation, Mathematica will conduct qualitative, in-depth interviews with implementers including the activity's design contractors, rehabilitation supervisors, and key ESIDA staff.	Process analysis
Data we expect to be collected by Georgia's National Assessment and Examination Center (NAEC)		
Assessments of student learning	We anticipate that assessments of lower-secondary and upper-secondary students' learning in math, science, and literacy will be administered using NAEC's computer-adaptive testing system. Assessments would be developed and piloted by NAEC with technical oversight from Mathematica to ensure test instruments adequately measure variation in student learning.	Student test scores in math, science, and literacy
Data to be collected by local survey firm procured by MCA-G		
Student survey	Survey data on student characteristics, recall-based measures of attendance (to be validated using site visits), perceived determinants of student attendance, perceptions of school building quality and safety, self-reported respiratory health, and perceptions of time on task during the school day.	Student attendance rates, particularly in winter months Students' time on task, including hours of instruction, measures of science laboratory use, and measures of recreational facility use Perceptions of school safety and health
Parent survey	Survey data on family demographics and socioeconomic characteristics, recall-based measures of student attendance, perceived determinants of student attendance, and perceptions of school building quality and safety.	Perceptions of school safety and health
Teacher survey	Survey data on teacher experience, demographic characteristics, certifications, perceptions of the quality and safety of school facilities, recall-based measures of time spent on instruction, and student attendance records.	Perceptions of school safety and health Experience of teaching staff; percentage of certified teachers Time on task, including hours of instruction, science laboratory use, and recreational facility use
School director survey	Survey data on school director operations and maintenance practices, average operations and maintenance expenditures, school facility usage, and student attendance records.	Student attendance rates, particularly in winter months Average expenditure on heating, lighting, and other operations/maintenance expenses Changes in maintenance and school management practices
Attendance	During site visits, the local data collection firm will directly measure attendance by (a) visually confirming the presence/absence of the study's student sample and (b) completing student headcounts for comparison against national administrative data recording the number of enrolled students. This will provide the most reliable attendance measure possible and is the gold standard for participation measurement. Additionally, by collecting this data, Mathematica can validate the other measures of attendance that will be collected; if other measures are reliable, those data might provide more detailed participation rate records than what can be observed directly.	Student attendance rates

Component	Description	Outcome
Ratings of school infrastructure	The local data collection firm will provide enumerators to visit all study schools and visually assess the quality of schools' infrastructure systems. Enumerators will collect data on classroom conditions related to heating, (for example, temperature, air quality from wood stoves), lighting, water, lavatory, and recreational facilities. The evaluation team and local data collection firm will consider ways to limit manipulation of school conditions during data collection, including unannounced visits and multiple visits per school year.	Measures of classroom conditions, including indoor temperature, air quality related to wood stoves, and adequate lighting Measure of overall building infrastructure quality
Qualitative research	In the second follow-up year, qualitative data collection will occur in a subsample of treatment and comparison schools. This will include: in-depth interviews with school directors to assess infrastructure usage patterns, community usage of the school building outside the school day, school operations, and maintenance practices; in-depth interviews with teachers to assess infrastructure usage patterns, time on task, and perceptions of school building quality and safety; and focus groups with students to assess infrastructure usage patterns, time on task, determinants of student attendance, and perceptions of school quality and safety.	Changes in maintenance and school management practices Perceptions of school safety and health Student and teacher time on task, including hours of instruction, science laboratory use, and recreational facility use

6. Analysis plan

We will estimate the impacts of the school rehabilitation activity using the following ordinary least squares regression:

(1) $$Y_{ist} = \alpha + \beta * TREAT_s + \boldsymbol{X_{is,t-1}} * \boldsymbol{\gamma} + \varepsilon_{ist}$$

where Y_{ist} is the outcome of interest (for example, test scores in science) for student *i* in school *s* measured at time *t*, which is either the first or second follow-up year in this case (in other words, impacts will be estimated separately for each outcome year). $TREAT_s$ is the treatment dummy variable indicating whether a school was randomly assigned to receive treatment; $X_{is,t-1}$ is a set of student-level demographic characteristics, baseline test scores, a set of school-level variables defining the random assignment blocks, and measures controlling for differences in the probability of treatment across random assignment blocks; and finally, ε_{ist} is the random error. The estimated value of the coefficient β represents the impact of the school rehabilitation program on the outcome of interest. Standard errors in the model will be clustered at the school-level using the standard Huber-White estimator to account for the possibility of correlations among individuals' characteristics within schools.

The study will also include subgroup analyses designed to measure whether there is a statistically significant difference between the magnitudes of programmatic impacts for key subgroups of students (relative to the impacts of the program among students outside each subgroup). Subgroup analyses will include disaggregated impact estimates based on gender, baseline test score levels, and measures of social vulnerability. In addition, we will examine the study's baseline survey data to investigate other potential subgroups of interest. In particular, if there is substantial variation before program implementation in baseline levels of school infrastructure quality, we will perform subgroup analyses to test whether the program was particularly effective in settings where building quality was especially low before program implementation.

The RCT-based estimates of the program's impact will also be used to estimate the activity's ERR and conduct beneficiary analyses. The ERR is a summary statistic that reflects the economic merits of the investment. Conceptually, it is the discount rate at which the cumulative benefits of an intervention over time are exactly equal to its costs; a higher (positive) ERR represents higher benefits and lower costs. According to the ERR documentation provided to Mathematica, MCC initially modeled the ERR for the IGEQ Project and produced an ex-ante ERR estimate of 11 percent, based on expected costs and benefits of the program's activities. After we complete impact analyses for this activity, we will conduct a similar exercise ex-post by comparing the activity's realized costs to evaluation-based estimates of the program's benefits. The exercise will enable MCC and other stakeholders to determine whether the project was a sound investment; it will also permit comparisons to other investments in Georgia. The accuracy of cost-benefit analyses depend on the plausibility of economic modeling assumptions and the precision of the impact estimates used to calculate program benefits over time. To address these issues, we plan to test the sensitivity of our ex-post ERR estimates to key parameters by using the confidence bounds of our impact estimates rather than point estimates.

The evaluation will also include an ex-post beneficiary analysis, which is an extension of the ERR that seeks to disaggregate income gains attributable to the investment for different segments of society. Such an analysis is critical to identifying the beneficiaries of the investment and determining if the activity is likely to lead to a reduction in poverty. We will conduct the beneficiary analysis separately by poverty category to determine the extent to which socially vulnerable or poor students reaped the benefits of the program's education investments. For each beneficiary group defined by poverty, we will determine the number of beneficiaries, the present value of benefits accruing to beneficiaries, and the cost-effectiveness of the investment (the present value of benefits per dollar spent).

For analyses of qualitative data, Mathematica will use qualitative transcript-coding software to organize and synthesize the key themes that emerge from document reviews, in-depth interviews, and focus groups. Where appropriate, we will compare and contrast from these data sources with descriptive data available in the study's quantitative surveys of students, teachers, parents, and school directors. These analyses will focus in particular on insights and themes that might play an explanatory role in understanding findings from the study's impact analyses. For example, if the impact analysis uncovered evidence of positive program impacts on some outcomes but not others, we would examine the study's qualitative data to develop a deeper understanding of the relationship between specific rehabilitation activities (such as constructing science laboratories in treatment schools) and the program's impacts.

7. Evaluation risks and monitoring plan

There are several risks to the study's internal and external validity that will require careful monitoring and management throughout the evaluation period. In particular, the program's selection process for determining whether treatment schools are ultimately rehabilitated will determine (1) whether the study's random assignment design is internally valid (unbiased); and (2) whether the evaluation includes enough rehabilitated schools to provide sufficient statistical power to detect the program's impacts.

At every stage of the program, Mathematica will remain in close contact with the MCA-G to monitor implementation of the school rehabilitation activity. Because our impact evaluation relies on a random assignment design, it will be particularly important to monitor the rehabilitation status of the schools assigned to treatment status or control status.

Several potential scenarios could lead to differences between the initial sample of schools assigned to treatment and the set of schools that ultimately receive rehabilitation. One major source of uncertainty is the cost of the rehabilitation activities relative to the program's budget. If the program cannot afford to rehabilitate all schools in the treatment group, we would help facilitate negotiations between MCC and MCA-G to develop a uniform approach to targeting the program to a subset of treatment group schools. For example, MCC might wish to target the program to a subset of schools using the same cost-effectiveness criteria used to identify the initial list of treatment schools, or it might be feasible to randomly select the subsample that remains in the program. Likewise, if the program is able to afford rehabilitating a larger number of schools beyond the initial treatment group, we will identify an additional number of schools from the eligible pool and randomly assign them to the treatment or control group.

We also anticipate that it might not be feasible to rehabilitate some of the schools assigned to treatment. For example, design assessments might uncover serious structural flaws or environmental hazards (meaning it would be more cost-effective to demolish and replace the building), or legal issues might arise such as unclear or disputed land titles. Ideally, such exclusions from the program would take place only due to uniform and readily identifiable exclusion rules that could be applied consistently in the evaluation's control group as well. For example, if after random assignment the program decided not to rehabilitate schools in a given region that had buildings larger than a certain size, Mathematica would consider excluding all treatment and all control schools meeting the exclusion criteria from the evaluation sample (this would preserve the equivalence of the treatment and control schools remaining in the evaluation sample).

However, in many instances we anticipate that the exclusion rules applied to treatment schools might be difficult or expensive to measure in control group schools. In particular, it is likely that some exclusion criteria would rely on measurements from detailed engineering assessments or expensive seismic tests, which would not be feasible to conduct across the full control group. In these cases, when MCC and MCA-G agree not to provide the program to a treatment school, that school would nonetheless remain in the evaluation sample as part of the treatment group. This will preserve the equivalence of the treatment and control group that was established by random assignment.

In instances where the program has funds to replace an excluded treatment school, we will facilitate a targeting process for possible replacement schools and randomly assign replacements to the treatment or control group. Specifically, in these instances MCC and MCA-G will work with Mathematica to adhere to the following procedure (as agreed in the "Pre-Randomization Memo" finalized between MCC and MCA-G in September 2014):

1. For each construction phase, MCA-G contractor(s) will conduct in-depth conditions assessments at all of the treatment schools in that phase, ensuring that all practical efforts are made to ensure rehabilitation of treatment schools wherever feasible.
2. Where previously agreed criteria merit exclusion of treatment schools, MCA-G will coordinate with the relevant contractors to submit a written justification to MCC and MCA-G proposing exclusion.

3. MCC will then consult with MCA-G to determine how many additional schools should be added to the treatment sample for that construction phase. (Depending on cost estimates for the schools in the phase, the number of replacement schools needed may be larger or smaller than the number of excluded schools.)
4. Upon agreement on the number of replacement schools, Mathematica will use the same targeting approach applied originally (preserving the specified regional allocation of program schools, and targeting schools with higher building utilization rates) to identify a 'replacement' pool of schools for that construction phase.
5. Mathematica will then randomly select half of the schools in the replacement pool to join the treatment group for that phase and the remainder will be added the evaluation's control group for that phase.

It is important to note that although the evaluation design does include plans for management of the school replacement process, excluding treatment schools will produce a marginal decline in the statistical power of the evaluation. In other words, dropping treatment schools from the program increases the likelihood that the impact evaluation will not be able to detect changes in important outcome variables (see Section 4.5 for additional details). The evaluation team will coordinate closely with MCC, MCA-G, and other program stakeholders to communicate the effects of various implementation decisions on the study's statistical power. Whenever possible, the number of excluded treatment schools should be minimized.

8. Administrative considerations

8.1. IRB requirements and clearances

Mathematica will prepare and submit an institutional review board (IRB) application for approval of the research and data collection plans. The application materials include three sets of documents: (1) a research protocol, which will draw heavily on the present design report and adds more information about plans for protecting study participants' confidentiality and human rights; (2) copies of all data collection instruments; and (3) a completed IRB questionnaire that summarizes the key elements of the research protocol, plans for protecting participants' human rights, and possible threats to participants if their confidentiality were compromised. Based on prior experiences, we expect that the study will qualify for expedited review because it presents minimal risk to participants. If so, the IRB can typically review the application within one week of its submission.

IRB approval is valid for one year from the date approval was granted and it must be renewed on an annual basis. We expect that the annual renewals will require minimal updates to the core application materials. Additionally, if data collection instruments change substantially from those that were approved by the IRB, then we must reapply for approval. Small changes to the instruments (such as rewording of questions, reordering of questions, or editing changes) do not require reapplication, but the finalized instruments must be submitted to the IRB for documentation.

After Mathematica drafts the IRB research protocol, we will coordinate with MCA-G to ensure the data collector and local stakeholders agree on the data collection protocol. Because Mathematica does not have a contractual relationship with the data collector, the data collector's contract with MCA-G must specify that they shall abide by the IRB's recommendations. The data collector and Mathematica must also sign an IRB authorization agreement stating that the data collector will adhere to the IRB-approved data collection procedures and protocols.

8.2. Data access, privacy, and documentation

After each of the baseline, interim, and final reports is produced, we will prepare corresponding de-identified data files, user manuals, and codebooks that may be made available to the public. These data files, user manuals, and codebooks will be de-identified according to the most recent guidelines set forth by MCC. The public use data files will be free of personal or geographic identifiers that would permit unassisted identification of individual respondents or their household, and we will remove or adjust variables that introduce reasonable risks of deductive disclosure of the identity of individual participants. Mathematica will remove all individual identifiers, including names, addresses, telephone numbers, government-issued identification numbers, and any other similar variables. We will also remove unique and rare data using local suppression, replacing these observations with missing values instead. If necessary, we will also use top/bottom coding, setting upper and lower bounds to remove outliers and collapse any variables that make an individual highly visible depending on geographic or other factors (such as ethnic classifications or languages spoken) into less easily identifiable categories. Finally, we will introduce random errors into any gathered geographic data (for example, global positioning system or geographic information system coordinates), displacing urban points 0 to 2 km and rural points 0 to 5 km, and additional 1 percent of rural points 0 to 10 km. Data perturbation will take place in a manner that will not significantly degrade the data.

8.3. Dissemination plan

Mathematica will present baseline and final evaluation findings in person to MCC and to stakeholders in Georgia. The timing of the analysis and reporting for the study will be determined by the program's phased rollout schedule. Thus, the baseline analysis will occur after data collection is completed for all three phases, following the 2016–2017 school year, and the analysis of the program's first-year impacts will occur following the 2017–2018 school year. We will work with MCC to increase the visibility of the project's findings, particularly among education policymakers and development practitioners. We will collaborate with MCC and stakeholders to identify a variety of forums—including conferences, workshops, and publications—to share results and encourage donors, implementers, and policymakers to integrate the findings into future programming.

For example, in addition to the project's full impact reports, we will develop short issue briefs summarizing and visualizing key findings for a broader audience of readers and stakeholders. Potential conferences for presenting evaluation findings will include forums hosted by the Comparative International Education Society, the American Evaluation Association, or the Association for Public Policy Analysis and Management. We will also seek to publish one or two peer-reviewed articles disseminating the study's results in journals such as the *Journal of Development Economics, American Economic Journal: Applied Economics,* or the *World Bank Economic Review.*

8.4. Evaluation team roles and responsibilities

Mathematica's project team has extensive experience conducting mixed-methods, multicomponent, large-scale evaluations in the field of education. **Mr. Matt Sloan** will serve as the program manager, acting as the primary point of contact for MCC. He will manage the relationships with government agencies and other local entities and contractors, while supervising the evaluation design and implementation process and ensuring high data quality. **Dr. Leigh Linden** is the evaluation's co-principal investigator, serving as senior analyst specializing in education evaluation and assessment and leading efforts to understand IGEQ impacts on learning. He will provide leadership on evaluation design and data analysis tasks. **Dr. Ken Fortson** is a co-principal investigator, providing methodological and technical oversight and support to the project team. **Mr. Ira Nichols-Barrer** will manage the quantitative data collection and lead implementation of the study's analysis tasks, and **Ms. Jessica Jacobson** will oversee the qualitative data collection and analysis process. **Ms. Natia Gorgadze** serves as the project's in-country consultant, providing substantive knowledge of Georgia's education system and assisting with the study's data collection and other local evaluation-management tasks.

8.5. Budget

At this time, Mathematica does not anticipate that the evaluation design and data analysis plans described in this report will require changes to the evaluation budget presented in the study's original proposal. Mathematica will work closely with MCC and MCA-G to ensure data collection is feasible within the compact's budget parameters.

REFERENCES

Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy,* vol. 2, no. 1, 2010, pp. 1–30.

Berry, Michael. "Healthy School Environment and Enhanced Educational Performance. The Case of Charles Young Elementary School." Washington, DC: Carpet and Rug Institute, January 2002.

Burde, Dana, and Leigh Linden. "Bringing Education to Afghan Girls: A Randomized Control Trial of Village-Based Schools." American Economic Journal: Applied Economics, vol. 5 no. 3, July 2013, pp. 27–40.

Cabezas, Veronica, Jose Cuesta, and Francisco Gallego. "Effects of Short-Term Tutoring on Cognitive and Non-Cognitive Skills: Evidence from a Randomized Evaluation in Chile." Working paper. Cambridge, MA: Jameel Poverty Action Lab, May 2011.

Deke, John, Lisa Dragoset, and Ravaris Moore. "Precision Gains from Publically Available School Proficiency Measures Compared to Study-Collected Test Scores in Education Cluster-Randomized Trials." Document No. PR 10-80. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, October 2010.

Dobbie, Will, and Roland G. Fryer, Jr. "Getting Beneath the Veil of Effective Schools: Evidence from New York City." *American Economic Journal: Applied Economics,* vol. 5, no. 4, 2013, pp. 28–60.

Dumitrescu, Anca, Dan Levy, Cara Orfield, and Matt Sloan. "Impact Evaluation of Niger's IMAGINE Program." Washington, DC: Mathematica Policy Research, September 2011.

Durán-Narucki, Valkiria. "School Building Condition, School Attendance, and Academic Achievement in New York City Public Schools: A Mediation Model." *Journal of Environmental Psychology,* vol. 28, no. 3, 2008, pp. 278–286.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts." Cambridge, MA: Mathematica Policy Research, January 2012.

Hedges, Larry V., and E. C. Hedberg. "Intraclass Correlation Values for Planning Group-Randomized Trials in Education." Education Evaluation and Policy Analysis, vol. 29, no. 1, 2007, pp. 60–87.

Hoxby, Caroline M., Sonali Murarka, and Jenny Kang. "How New York City's Charter Schools Affect Student Achievement: August 2009 Report." Cambridge, MA: New York City Charter Schools Evaluation Project, September 2009.

Levy, Dan, Matt Sloan, Leigh Linden, and Harounan Kazianga. "Impact Evaluation of Burkina Faso's BRIGHT Program." Washington, DC: Mathematica Policy Research, June 2009.

Muralidharan, Karthik, and Venkatesh Sundararaman. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *The Economic Journal*, vol. 120, no. 546, 2010, pp. F187–F203.

Nichols-Barrer, Ira, Matt Sloan, Ken Fortson, and Leigh Linden. "Program Logic Assessment for the Georgia Improving General Education Quality Project." Draft report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, December 2013.

Woolner, Pamela, Elaine Hall, Steve Higgins, Caroline McCaughey, and Kate Wall. "A Sound Foundation? What We Know About the Impact of Environments on Learning and the Implications for Building Schools for the Future." *Oxford Review of Education,* vol. 33, no. 1, 2007, pp. 47–70.

Yeh, S.S. "The Cost-Effectiveness of Five Approaches for Raising Student Achievement." *American Journal of Evaluation,* vol. 28, no. 4, 2007, pp. 416–436.