Evaluation of the Georgia II Improving General Education Quality Project

Final Report

June 13, 2023

Ira Nichols-Barrer, Ivonne Padilla, Elena Moroz, and Matt Sloan

Submitted to:

Millennium Challenge Corporation
1099 Fourteenth St., NW, Suite 700
Washington, DC, 20005
Contract Number: MCC-13-BPA-0040 (CL-002)

Submitted by:

Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763
Project Director: Matt Sloan
Reference Number: 40306.700

This page has been left blank for double-sided copying.

Acknowledgements

This report reflects the contributions of many people. We are grateful to our Millennium Challenge Corporation (MCC) project monitor, Jenny Heintz, and contracting officer's representative, Ryan Moore, for their thoughtful guidance and support throughout the evaluation. During the compact, MCC's in-country team of Jenner Edelman, Sonia Shahrigian, and Marina Kutateladze also provided extensive and valuable support and insights.

We also acknowledge the wide range of staff from the Millennium Challenge Account-Georgia and its successor entity the Millennium Foundation, led by Magda Magradze, as well as project implementers who generously shared their time and attention to help improve the quality, comprehensiveness, and depth of the study. We are grateful to Government of Georgia staff at the Ministry of Education, the Teacher Professional Development Center, and the Educational and Scientific Infrastructure Development Agency for generously sharing their time and expertise and providing important information about the project's expected outcomes, implementation, and activities. We also received indispensable support and advice from staff of the Millennium Challenge Account-Georgia office, especially Zura Simonia, who managed the evaluation's data collector during the compact and provided both substantive and technical expertise at every stage of the study, Nino Udzilauri, who guided evaluation planning discussions and facilitated a wide range of data collection efforts for the project's teacher and school director training activities, and Kartlos Kipiani, who generously shared planning and implementation details about the school rehabilitation activity.

This report also depended on contributions from many data collection, supervisory, and support staff. We are grateful to the staff of the Institute for Polling and Marketing (IPM) and the National Assessment and Examination Center (NAEC) for carrying out the study's extensive survey, qualitative data collection, and learning assessment activities. We also wish to thank the many students, parents, teachers, and school directors who responded to our surveys and participated in in-depth interviews. Many components of the study's data collection would not have been possible without the contributions of Natia Gorgadze, who supervised data collection activities as a locally based member of Mathematica's evaluation team. At Mathematica, Larissa Campuzano provided excellent comments on the draft report. We also thank Susan Gonzalez and Ariel Mendez for editing services, and Sharon Clark for production support.

Mathematica strives to improve public well-being by bringing the highest standards of quality, objectivity, and excellence to bear on the provision of information collection and analysis to our clients. The findings in this report solely reflect Mathematica's interpretation of available information. Mathematica staff involved in analyzing the information and authoring this report did not report any conflicts of interest; the evaluation was funded exclusively by MCC.

This page has been left blank for double-sided copying.

Contents

Tables

Figures

This page has been left blank for double-sided copying.

Executive Summary

A. Overview of the Georgia Compact and interventions evaluated

Recognizing that economic growth in Georgia is being significantly hindered by skills shortages and education gaps in the workforce, the Government of Georgia and the Millennium Challenge Corporation (MCC) carried out the $139 million Georgia II Compact to improve the quality of education in science, technology, engineering, and math (STEM), and thereby develop a more skilled Georgian labor force. The five-year compact, which entered into force in July 2014 and concluded in July 2019, included three projects that focus on general education, workforce development, and higher education. This report presents final evaluation results for the compact's general education project.

The Improving General Education Quality (IGEQ) Project invested a total of $71 million, allocated across three project activities. The largest of these was the Improved Learning Environment Infrastructure (ILEI) activity, which invested in school rehabilitation to provide safe learning environments that include adequate classroom facilities, improved science labs, functional central heating, and improved sanitary facilities. Second, the Training Educators for Excellence (TEE) activity supported professional development by training and mentoring teachers in subjects related to science and math and by training principals to strengthen school management. The Project also included a smaller Education Assessment Support activity designed to support rigorous learning assessments and a results-oriented education system.

B. Evaluation type, questions, and methodology

MCC contracted with Mathematica to evaluate the implementation and potential effects of the IGEQ project, with a focus on the ILEI school rehabilitation activity and the TEE activity. The study's interim evaluation report assessed if these activities were implemented as designed during the compact, and examined if the proximate outcomes of each activity among beneficiaries was consistent with the Project's theory of change (Nichols-Barrer et al. 2019). This final report extends the analysis and examines the impacts of these investments, including multiple years of data collected after the compact came to an end.

For the school rehabilitation activity, the evaluation carried out an impact evaluation using a randomized controlled trial to estimate the impact of these investments on school infrastructure, the learning environment, school operations and management, and student learning outcomes. For this final report, the evaluation's research questions for this activity are:

1. 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 students' and teachers' perceptions of health and safety?
2. What are the impacts of the ILEI activity on teachers' behavior, such as attendance and time spent teaching?
3. What were the impacts of the ILEI activity on students' outcomes? What are the impacts on attendance, enrollment, dropout and retention rates, time spent studying in and out of school, and learning outcomes?
4. 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)?

For the TEE activity, the evaluation carried out a mixed-methods performance evaluation assessing how teachers and school directors responded to the activity's training sequence. For this final report, the research questions for the TEE activity are:

1. Did teacher training modules improve teachers' knowledge about student-centered instruction, formative assessments, and classroom management?
2. Did teacher training modules improve teachers' willingness to use student-centered instruction, formative assessments, and classroom management?

C. Implementation summary

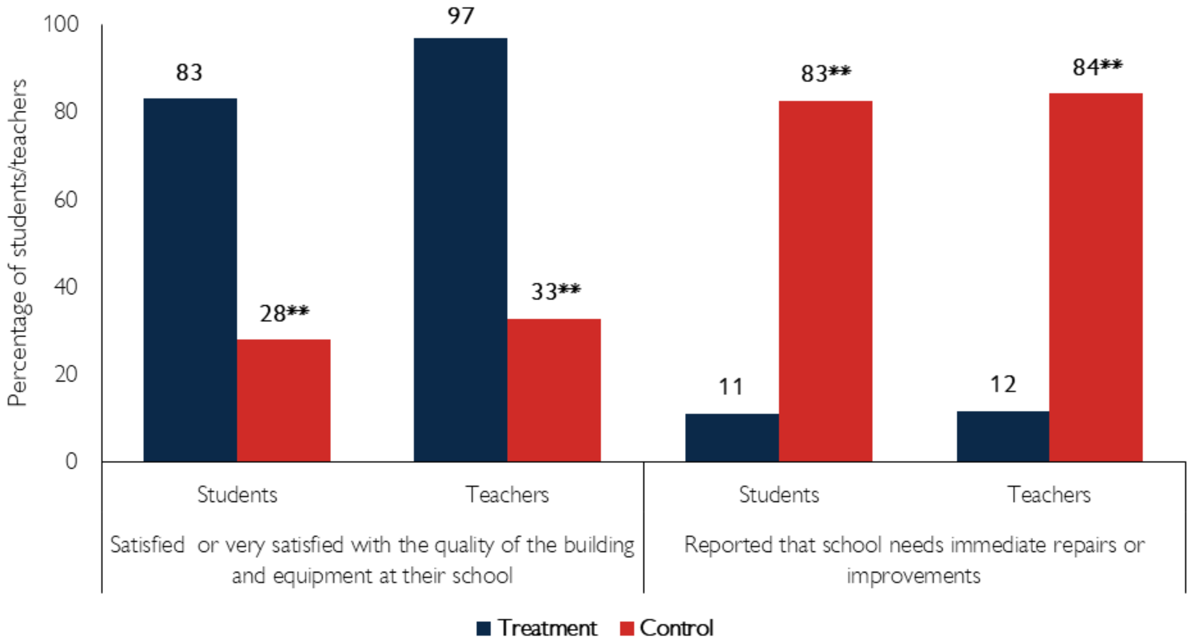
This study's interim evaluation report (Nichols-Barrer et al. 2019) found that the project had been implemented in a manner that closely aligned with the project's original theory of change, meeting most of the project's implementation and output targets. The status of project implementation at the end of the compact was as follows:

- **ILEI school rehabilitation activity:** This activity completed rehabilitation of 91 school buildings serving grades 7-12, with investments targeted to such problems as inadequate heating systems, poor indoor air quality, and inadequate lighting. Each rehabilitated school received interior and exterior rehabilitation of classroom flooring, walls, and windows; new or significant upgrades to utilities such as electricity, heating, water, and wastewater; and each school also received a new or upgraded science laboratory classroom. The Georgian government and implementers identified schools to be targeted for rehabilitation using a formula that prioritized schools outside of the capital Tbilisi according to their physical condition (dilapidated physical infrastructure), social vulnerability (higher proportion of socially vulnerable students), number of students enrolled, and the ratio of each building's size to the number of enrolled students (Georgia II Star Report, 2021).
- **TEE activity:** This nationwide training initiative sought to improve grade 7-12 instruction in math, science, geography, and English (by delivering several training modules to teachers in these subjects over the course of a one-year training sequence) and improve school management (by delivering a more extensive series of training modules to school directors, over the course of a two-year training sequence). Ultimately, during the compact the TEE activity offered training to all 2,085 public school directors in Georgia, at least one professional development facilitator in each of these schools, and all of the country's 18,750 teachers delivering grade 7-12 instruction in the targeted subjects of math, science, geography, and English (Georgia II Star Report, 2021).

D. Findings for the school rehabilitation activity

The school rehabilitation activity dramatically improved the quality of rehabilitated buildings. Findings from the evaluation's randomized control trial reveal that the activity had a consistent and large impacts on a wide range of infrastructure outcomes (Figure ES.1). Some of the key upgrades in rehabilitated schools included eliminating widespread problems with classroom walls, ceilings, and floors, installing electrical lighting systems in classrooms that had no functional lighting before, delivering central heating systems that improved classroom temperatures and eliminated serious air quality problems associated with wood-stoves, and upgrading sanitary facilities (with running water and flush toilets) and science labs (with lab benches and equipment for experiments and demonstrations) in highly visible ways that were noticed and appreciated by students and teachers. While school directors reported an increase in operating costs associated with these investments (particularly utility costs related to using a central heating system), there is little evidence these costs were unmanageable. In some cases, the heating costs in rehabilitated schools were offset by revenues from increased student enrollment (particularly in early grades), as more families chose to enroll in rehabilitated school instead of other regional schools that did not receive the program.

Figure ES.1. Impact of rehabilitation on student/teacher perceptions of building quality



Source: Teacher and student surveys completed by 1,380 teachers and 8,460 students, interviewed at two-year follow-up.

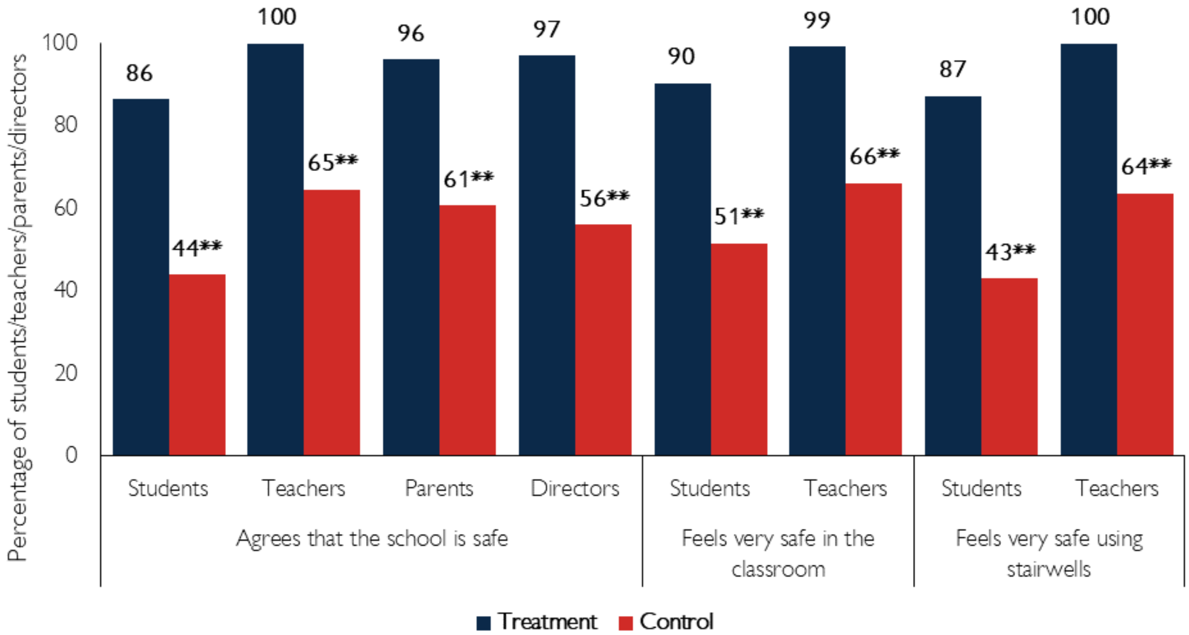
Notes: Outside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

Teachers reported that these infrastructure upgrades addressed multiple serious problems affecting classroom instruction. In surveys and in-depth qualitative interviews, teachers consistently reported that these infrastructure upgrades directly improved their ability to focus on instruction in the classroom. Functional electric lighting made it possible for students to read written materials more easily, indoor sanitary facilities helped students reach their classrooms on time, and in winter teachers reported that central heating brought their classrooms to a more-comfortable temperature (meaning students no longer needed to wear winter jackets during lessons) and addressed serious air quality problems related to poor ventilation and smoke from wood-burning stoves. While these changes did not affect teachers' overall attendance rates or the total number of minutes teachers spent teaching in the classroom each day, respondents consistently pointed out that the activity made it possible to focus more fully on instruction without the discomforts and distractions present in the building prior to rehabilitation.

Students agreed that infrastructure upgrades addressed serious barriers to learning in the classroom. In surveys and focus groups, data from students in rehabilitated schools revealed a striking pattern of improvements in views about their ability to focus on learning in the classroom. In schools that were not rehabilitated (the evaluation's control group) students pointed out that poor lighting, inadequate heating, harmful air quality, ceiling leaks, the lack of indoor toilets, and the absence of science labs were all serious problems affecting their comfort and ability to focus on instruction. None of these issues remained in rehabilitated schools. In addition to enhancing student comfort and ability to focus, rehabilitation also directly reduced interruptions during the school day (for example, the need to pause lessons to refuel stoves and change rooms to air out classrooms when wood smoke became overwhelming). That said, there is little evidence that school rehabilitation changed the total number of days students spend at school each year: the program did not have an impact on student absenteeism or dropout rates. Rather, students reported that the activity substantially improved the quality of the time spent on learning activities during the school day.

Figure ES.2. Impact of rehabilitation on student/teacher perceptions of safety



Notes: Outside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

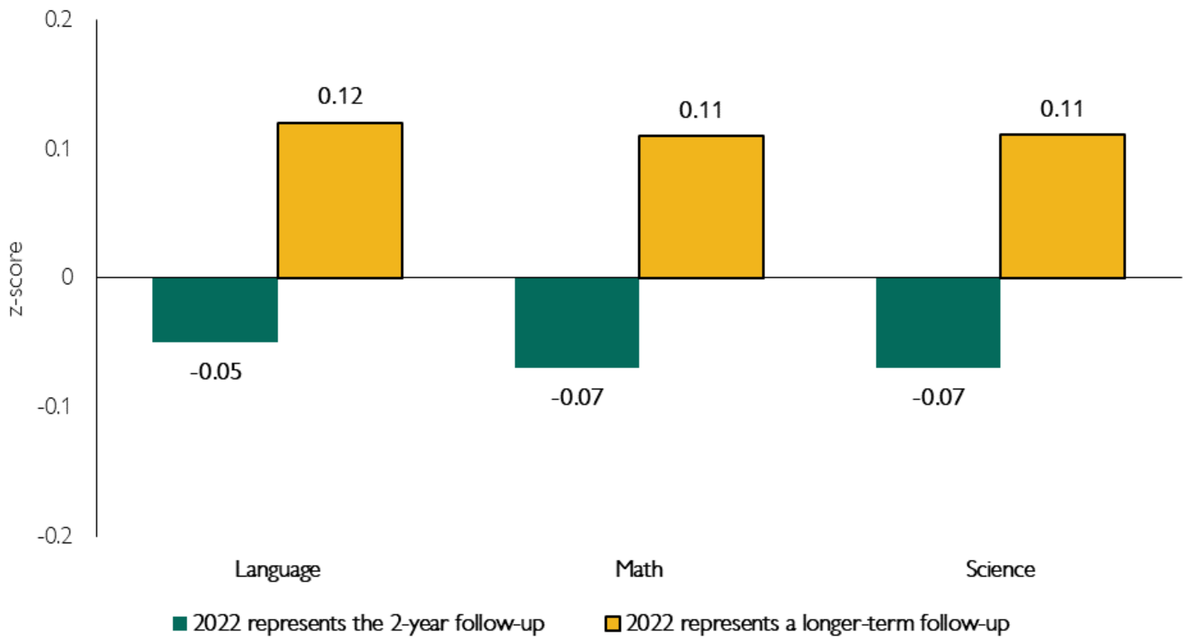
However, teachers reported difficulties with accessing and using upgraded science labs.

Rehabilitated school buildings received a single upgraded science lab which was designed to be shared across secondary grade-levels and with equipment spanning multiple subjects (including biology, chemistry, and physics). These upgraded science labs are being used actively—for example, rehabilitation substantially improved the likelihood that students will see a science demonstration or experiment during the school year. That said, science teachers in rehabilitated schools consistently reported that they have not been able to use the labs as often as they had hoped, and there have been barriers to integrating the new facilities into their curriculum and daily lessons. Scheduling access to the labs is difficult due to the number of grade-levels and subjects seeking to access the same facility, and in some schools the amount of equipment-storage needed in the labs also impeded airflow from the central heating system and made the labs to be uncomfortably cold in winter. These frustrations were exacerbated by requirements to enforce social-distancing regulations during the COVID-19 pandemic: since the layout of these labs are organized around conducting small-group experiments at shared tables, in many schools it was not possible to access the labs at all during the 2020-2021 or 2021-2022 school years.

The impact of rehabilitation on learning outcomes was ambiguous, in part due to the disruptions of the COVID-19 pandemic. After two years of access to rehabilitated schools, we did not find evidence that the Activity had an impact on math, language, or science test scores. An important complicating factor in this analysis is that the school closures caused by the COVID-19 pandemic occurred approximately halfway through the follow-up period in this evaluation: some schools in the sample completed the study's two-year follow-up period before the pandemic began, whereas other schools experienced extended disruptions and school closures during the evaluation's two-year follow-up period. While the study was not designed to carry out precise subgroup analyses testing for differences between these two groups of schools, the results suggest that learning outcomes differed dramatically during the pandemic. Particularly in language and math, rehabilitation appears to have had a positive effect on learning outcomes before the pandemic, but these effects were absent for schools assessed after the pandemic. If the pre-pandemic impacts of rehabilitation had held for the entire sample, impacts of that magnitude would have represented a meaningful amount of learning growth. For the language exam, the pre-pandemic impact of 0.13 standard deviations is equivalent to about 7 months of learning for upper-secondary students. Similarly, pre-pandemic impact on math scores (0.10 standard deviations) is equivalent to 5 months of learning. On the other hand, in science, the pre-pandemic effect of rehabilitation was close to zero—an outcome which might be explained by the barriers to science-lab access and use discussed above.

Three to five years after rehabilitation, schools have been able to maintain improved infrastructure and there is evidence of longer-run improvements in learning outcomes. Longer-term data collection activities conducted after the compact ended (between three and five years after rehabilitation was completed) show that schools have been able to maintain infrastructure improvements over time. Rehabilitation continued to produce a dramatic improvement in overall building quality, heating systems and heating-system usage, air quality, and sanitation facilities. Interestingly, these longer-term analyses also provide suggestive evidence that rehabilitation may produce more positive learning benefits beyond the second follow-up year. In all three subjects (language, math, and science), impacts on learning outcomes were negative or close to zero for schools in their second follow-up year, but impacts became positive (and larger in magnitude) in schools that were in their third, fourth, or fifth follow-up year (Figure ES.3). Importantly, these effects are not statistically significant, and the impact estimates are imprecise due to the limited number of schools where longer-term effects could be measured in time for this report. Nonetheless, it is noteworthy that longer-term impacts of this magnitude (0.12 s.d. in language, 0.11 s.d. in math, and 0.11 s.d. in science) would represent an educationally meaningful boost in student achievement. For example, an effect size of 0.11 standard deviations would be roughly equivalent to 5 months of learning in language and 7 months of learning in math, for students in grade 10.

Figure ES.3. Impact of rehabilitation on student test scores, by follow-up year



Source: Student learning assessments administered by the National Assessment and Examination Center (NAEC).

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in appendix B

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

E. Findings for the TEE activity

This study's interim evaluation report showed that the TEE activity succeeded in implementing the program on a nationwide basis. Successfully implementing a training intervention on such a scale was a remarkable achievement: the training sequence was offered to Georgia's entire population of school directors (2,085 individuals) and all of Georgia's upper-grade teachers in the subjects of science, mathematics, English, and geography (18,750 teachers in total). In terms of the training's potential effects, the interim analysis also found a consistent pattern of improvements in teachers' self-reported knowledge of student-centered instruction strategies in the initial period after training. However, outside of professional development activities (where we found a stronger pattern of improvements), the interim analysis did not reveal consistent evidence of short-term changes in teachers' classroom practices.

For the TEE activity, it is important to remember that the Activity's theory of change did not assume that teaching practices would improve immediately following completion of the training sequence. Instead, the program was designed to produce rapid improvements in teachers' knowledge and their professional development resources (through teacher study groups and other professional networks), which would in turn produce changes in their teaching practices and ultimately improve students' learning outcomes over longer periods of time. To examine whether this pattern occurred, this final evaluation report conducted a longer-term follow-up analysis of teachers' practices up to two years after the training sequence was completed.

The key findings from the TEE evaluation in this final report are as follows.

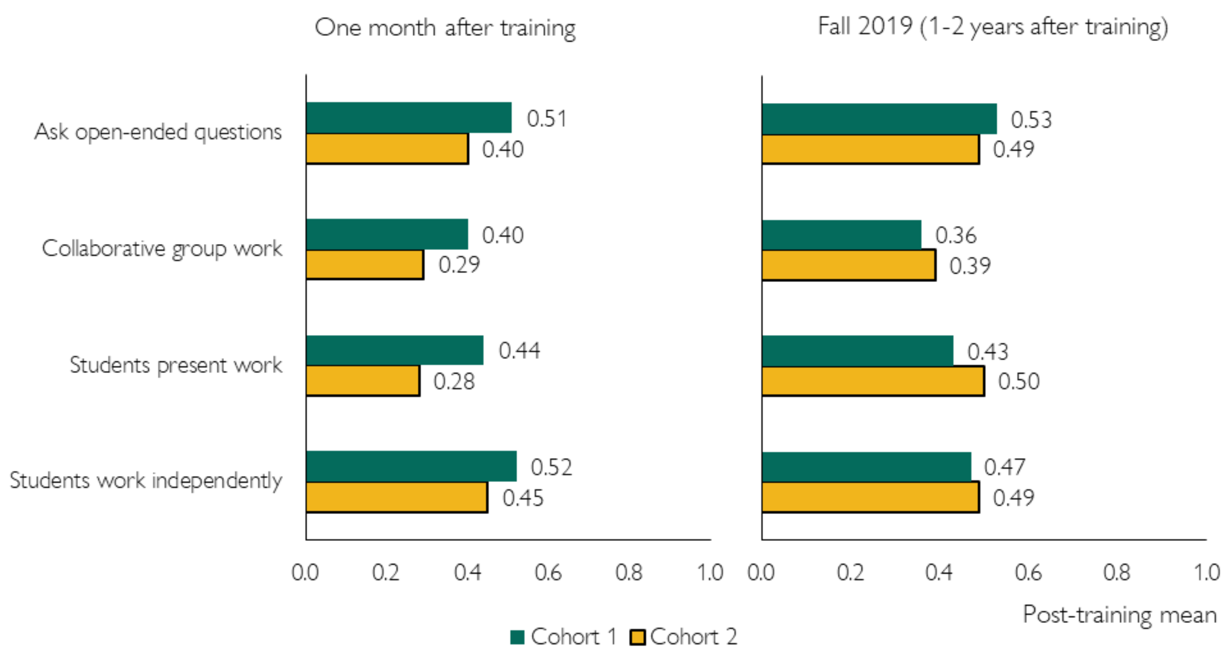
The Georgian government's 2019 incentivized retirement policy produced a teaching workforce that is younger, more likely to have completed the TEE training sequence, and more likely to be using TEE-supported teaching practices. In May 2019, the Ministry of Education announced a new retirement incentive program: teachers who had reached full retirement age were given a generous retirement incentive, and any eligible teachers who refused the offer would be required to pass a professional-competency exam to continue with their public-school careers. The goal of the policy was to encourage older teachers who might be less willing to undergo new types of professional development to retire and make space for younger teachers to enter the classroom. Our analyses show that the policy appears to have worked exactly as designed, and directly supported the goals of the TEE activity. These incentivized retirees comprised 12 percent of the evaluation's sample, and were (on average) 20 years older than non-retirees, 27 percentage points less likely to have completed the core TEE training modules, and substantially less likely report using TEE-related instructional practices such as carrying out daily informal assessments, asking open-ended questions, working with struggling students on a separate basis, or discussing inclusion of students with different ethnicities, religions, or sexual identities.

Two years after the TEE training sequence, nearly all teachers continued to report that they are confident or very confident in using the types of student-centered instruction practices that were the focus of the TEE activity. As with the interim evaluation results, the endline survey revealed that teachers remained highly confident that they had acquired enough knowledge to apply the types of teaching practices that were the focus of the TEE training sequence (although the evaluation did not collect direct observational data about use of these practices for the final evaluation report). There is very little evidence that the gains in confidence reported by teachers shortly after the training sequence ended have faded over time.

Among the first cohort of trainees (which prioritized teachers who had passed Georgia's teacher certification exam), use of TEE-supported practices in science instruction increased over time, but there were only modest changes in the use of other student-centered teaching practices. Over a two-year period after the training sequence ended for the first cohort of trainees (which largely consisted of more highly qualified "senior," "lead," or "mentor" teachers), there were only modest changes in teachers' self-reported use of student-centered teaching practices. These changes ranged from a decline of 5 percentage points (for example, in the use of daily lesson plans designed to achieve specific learning goals) to an increase of 4 percentage points (for making daily changes in instruction based on testing). However, there was a notably larger improvement in science-related teaching practices: during the two years after the training sequence ended, monthly use of lab experiments increased by 7 percentage points and monthly practice of hypothesis-testing increased by 8 percentage points.

However, among the second cohort of trainees (which prioritized less qualified teachers) there were large improvements in the use of teaching practices expected to improve students' critical thinking and collaboration, such as asking open-ended questions and having students present their work. Shortly after the training sequence ended, the second training cohort (which consisted of less qualified "practitioner" teachers who had not passed Georgia's teacher certification exam) reported using student-centered practices less often than the first cohort (which mostly consisted of teachers who had passed the certification exam). One year later (in fall 2019), teachers in the second cohort appear to have caught up with the first cohort in their use of certain TEE-supported instructional practices. For example, one year after the training sequence ended Cohort 2 teachers reported statistically significant improvements in their use of open-ended questions (an increase of 9 percentage points), collaborative group work (an increase of 10 percentage points), and having students present their work (an increase of 22 percentage points). Immediately after training, teachers in the first cohort (who were more qualified on average) were outperforming the teachers in the second cohort by a magnitude of 7 to 16 percentage points across these practices. By the endline survey in fall 2019, Cohort 2 teachers were only underperforming Cohort 1 teachers in their use of open-ended questions (by a magnitude of 4 percentage points), and they were outperforming Cohort 1 teachers on the other three practices (by a magnitude of 2 to 7 percentage points). These trends suggest that the TEE activity may be succeeding in producing longer-run changes in teaching practices, and doing so particularly among less-qualified teachers, as sought by the program.

Figure ES.4. Cohort 2 teachers' growing use of TEE-related teaching practices



I. Introduction

Recognizing that economic growth in Georgia is being significantly hindered by skills shortages and education gaps in the workforce, the Government of Georgia and the Millennium Challenge Corporation (MCC) carried out the $139 million Georgia II Compact to improve the quality of education in science, technology, engineering, and math (STEM) and thereby develop a more skilled Georgian labor force.[1] The five-year Compact, which entered into force in July 2014 and concluded in July 2019, included projects that focused on general education, workforce development, and higher education.

This report presents final evaluation results for the Compact's Improving General Education Quality (IGEQ) Project, with a total investment of $71 million allocated across three project activities. The largest of these was the Improved Learning Environment Infrastructure (ILEI) activity, which invested in school rehabilitation to provide safe learning environments that include adequate classroom facilities, improved science labs, functional central heating, and improved sanitary facilities. Second, the Training Educators for Excellence (TEE) activity supported professional development by training and mentoring teachers in subjects related to science and math and by training principals to strengthen school management. The project also included a smaller Education Assessment Support activity designed to support rigorous learning assessments and a results-oriented education system. MCC chose Mathematica to conduct a rigorous evaluation of the ILEI and TEE activities and examine their impacts on both intermediate and long-term outcomes.

The evaluation involves a mixed-methods approach that draws on both qualitative and quantitative data to explore how the project was implemented and its ultimate effects on the Georgian education system. In this report, we present the final results of the evaluation using data collected between 2017 and 2022, which includes several years of follow-up data collected after the Compact came to an end in 2019. The evaluation draws on a rich array of data collection components including technical building infrastructure assessments; surveys of school directors, teachers, students, and parents; student learning assessments; and in-depth qualitative interviews and focus groups with school directors, teachers, and students. To provide context, we next describe the IGEQ Project's activities and logic model and briefly review the existing literature on the impacts of similar infrastructure improvement and teacher professional development programs in other developing countries. We then summarize the objectives of the final report.

A. Overview of the IGEQ activities evaluated in this study

The objective of the IGEQ Project was to improve education quality by targeting the physical learning environment, secondary school teacher subject knowledge and pedagogical skills, school management, and education assessments, with an emphasis on the STEM subjects. This evaluation assesses two different activities: (1) a school rehabilitation activity focused on intensive infrastructure investments in a subset of schools and (2) a nationwide teacher and school director training initiative designed to improve the capacities of secondary school educators in fields related to STEM throughout Georgia. The teacher training activity focused on educators in grades 7 to 12 in the subjects of science, mathematics, English, and geography.

[1] All project costs reported here are based on final Compact disbursement figures from MCC, as recorded in the Georgia II Compact's STAR Report.

The school rehabilitation activity aimed to upgrade the quality of physical infrastructure and create an improved learning environment in rehabilitated schools. Examples of rehabilitation projects were:

- Systems for heating (replacing wood stoves with central heating)
- Lighting
- Water and plumbing
- Sanitary facilities
- Recreational facilities
- Science laboratories
- Building interiors (flooring, stairs, and classroom walls)
- Building exteriors (roofing and masonry)

Through a random assignment process, the Activity selected 104 schools throughout Georgia to receive detailed rehabilitation designs. Where rehabilitation was feasible, work in these schools took place over the course of several construction seasons (the 2015–2016 school year, the 2016–2017 school year, the 2017–2018 school year, and the 2018–2019 school year).

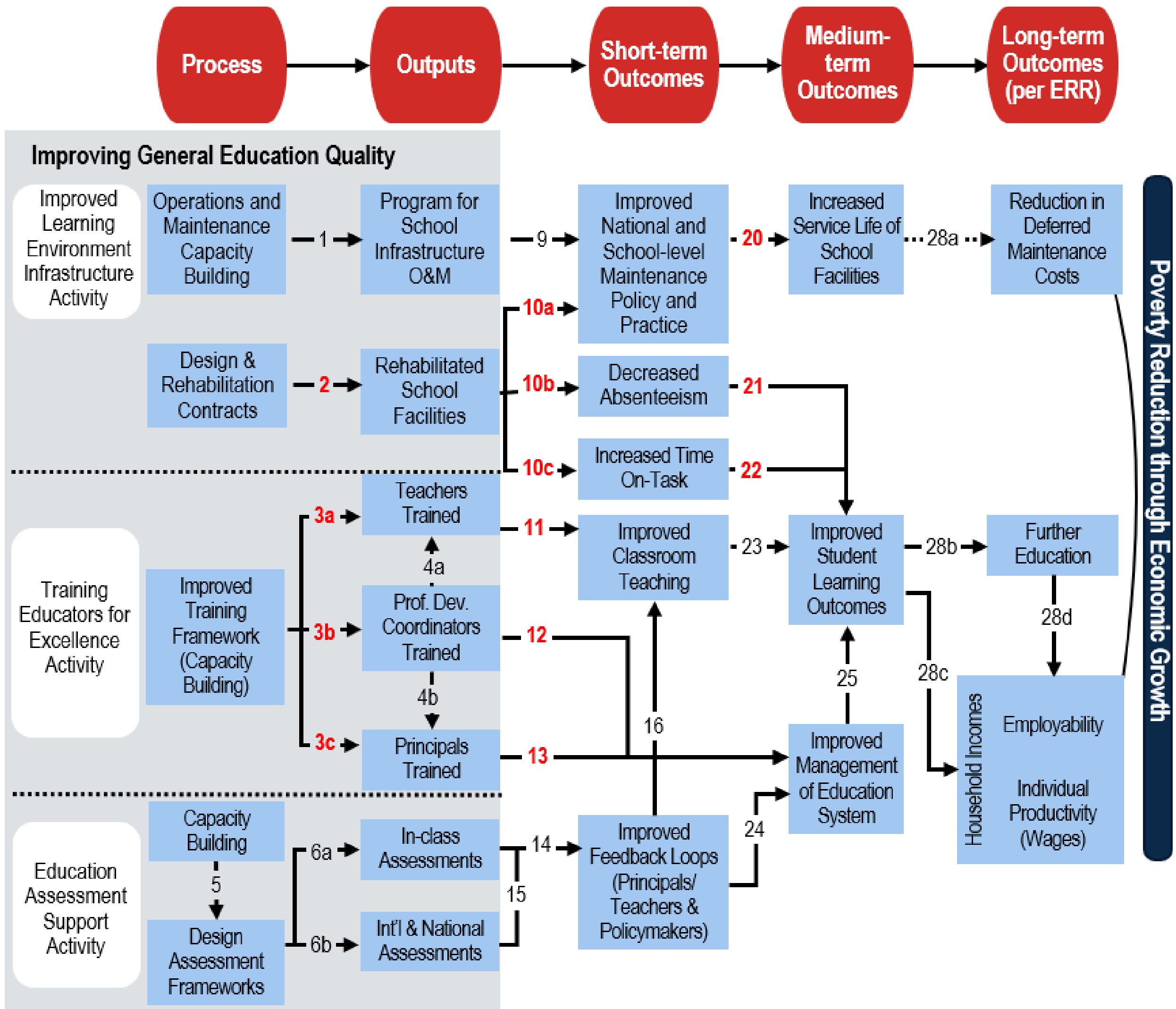
The one-year teacher training sequence delivered under the TEE activity was broken up into three core modules, plus an additional subject module specific to the teacher's primary teaching subject (mathematics, science, English, and geography). Each module involved an in-person training session lasting between two and five days. The three core modules for teachers covered the characteristics of a student-centered learning environment (encouraging differentiated instruction and opportunities for critical thinking and creativity); instructional and assessment strategies (lesson planning with learning objectives, and using ongoing formative assessments alongside summative assessments); and classroom management and teacher professional practice (encouraging use of collaborative group work, and encouraging teachers to engage with professional networks and teacher study groups). The TEE training sequence for school directors (the Leadership Academy) was delivered over the course of two years in a series of five modules addressing instructional leadership practices, staff management skills, and training in financial management related to directors' oversight of school budgets.

According to the program's logic model (Figure I.1), the inputs (rehabilitating the schools and training teachers and school directors) are intended to improve students' learning outcomes, but the mechanisms for improving learning differ. In the case of school rehabilitation, the intervention aims 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 that rehabilitating schools could plausibly improve the health and well-being of students, which might provide another pathway for the intervention to affect learning and other long-term outcomes.[2] In the case of the training activities for teachers and school directors, the program is intended to improve students' achievement outcomes by directly improving the quality of

[2] Children might also be exposed to poor air quality and sanitation at home, meaning that rehabilitating schools is unlikely to remove all the health risks that students face. Because treatment was assigned randomly in this evaluation, we can expect home air quality and sanitation to be equivalent between treatment- and control-group homes. Thus, this study can attribute any health improvements observed in the treatment group to the school rehabilitation intervention.

classroom instruction (through improved teaching practices) and school management (including instructional leadership from school directors).

Figure I.1. The IGEQ program logic



Source: MCC Georgia II Compact, Annex II.

Note: Arrows with dotted lines refer to links that MCC expects cannot be evaluated or measured. Links are uniquely numbered (e.g., "1," "2," "3a," "3b," "3c"). ERR = economic rate of return; IGEQ = Improving General Education Quality; MCC = Millennium Challenge Corporation; O&M = operations and maintenance expense.

The program logic developed by MCC and the Georgian government's implementation body for the Compact, Millennium Challenge Account-Georgia (MCA-G), 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 reducing poverty through economic growth. Each of the links 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.

Before the evaluation began, we assessed the plausibility of the IGEQ program logic and the associated ERR calculations MCC developed prior to the project. To do so, the evaluation team reviewed the available evidence on the impacts of similar program designs in other contexts and discussed it extensively with local education experts and IGEQ stakeholders. These discussions included MCA-G staff, stakeholders in relevant GoG centers and ministries, and school staff interviewed during the team's site visits to schools selected for the ILEI rehabilitation program. We examined the program logic for each of the IGEQ components separately, reviewing the relevant literature on the effects of similar interventions in other contexts. We explained how an evaluation would contribute to addressing gaps in the literature and noted potential concerns about areas in which assumptions in the logic model might not hold (Nichols-Barrer et al. 2013). For example, the CBA and logic model for the school rehabilitation activity largely focused on educational attainment (additional years of secondary and postsecondary schooling) as the primary channel for improvements in long-term employment and wages; our assessment of the program logic noted that improved school infrastructure could have meaningful effects on student health (which was not part of the program logic at all) and student learning (which does appear in the program logic but was more narrowly linked to the provision of science labs in the initial CBA model), and explained that each of these outcomes could independently lead to long-term improvements in employment and earnings.

The various components of the evaluation's final data collection efforts and analyses in this report are all designed to provide information about the inputs in the logic model and the potential relationships between these inputs and evaluation outcomes. Specifically, the evaluation was designed to assess a subset of the assumed links in the logic model (highlighted in red in Figure I.1), on a separate basis for the school rehabilitation activity and the TEE activity. Taken together, these findings provide evidence on whether the ultimate goals envisioned in the program logic are likely to be realized. Next, we summarize our review of the relevant literature.

B. Literature review

An extensive area of academic literature investigates the relationship between educational inputs and measures of student learning, educational attainment, and employment outcomes. But much less is known about the effects of these interventions in developing countries, and little empirical work exists on the education system in Georgia. In our view, the existing evidence base does not support strong predictions about the size of the program's expected impacts for either the school rehabilitation or the training activities for teachers and school directors. We summarize the relevant literature here.

1. Prior evidence on school rehabilitation

According to the initial ERR calculations used for the school rehabilitation activity, MCC aimed for this intervention to produce the following improvements in students' long-term outcomes: a 10 percent improvement in the percentage of students transitioning into upper secondary school and a 10 percent improvement in the percentage of students transitioning into postsecondary programs. At the end of the Compact, MCC adjusted the ERR calculation and included an assumed increase in student learning as an additional parameter in the cost-benefit model. The evidence from prior studies shows great uncertainty regarding the relationship between school infrastructure inputs and all the aforementioned outcomes. Some evaluations of school construction and rehabilitation activities 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; Bagby et al. 2014; Bagby et al. 2017) and limited to no short-term impact in other contexts (Dumitrescu et al. 2011). Very little rigorous research assesses whether a causal link exists between school rehabilitation inputs and long-run improvements in employment rates or income levels; in fact, we are not aware of any studies that 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.

Past studies on school infrastructure have largely focused on, among other things, 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. But the impacts of infrastructure improvements likely depend on existing conditions in the affected facilities or communities. For example, if a program improves a school that is already functioning well, one expects the benefits of the program to be relatively modest. Conversely, in a community with very limited school facilities, construction or rehabilitation can at least potentially produce large benefits. As shown in this evaluation's baseline report (Nichols-Barrer et al. 2017), the schools selected for the rehabilitation activity in Georgia suffered from a wide range of infrastructure problems, suggesting that the activity had the potential to meaningfully improve the learning environment.

For example, impact evaluations of the Burkinabé Response to Improve Girl's Chances to Succeed (BRIGHT) program in Burkina Faso, an initiative that constructed and later expanded primary schools in 132 rural villages throughout the 10 provinces with the lowest rates of school enrollment for girls, specifically targeted communities that did not previously have ready access to a school. The evaluations found that BRIGHT schools had a positive impact on school enrollment and a large impact on test scores, primarily driven by large improvements in grade attainment (Levy et al. 2009; Kazianga et al. 2013; Davis et al. 2016). Several descriptive studies of school conditions in the United States 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 might improve student attendance, but not necessarily in the short-term. The Improve the Education of Girls in Niger (IMAGINE) program constructed schools in 10 communities with low enrollment and primary school completion rates for girls, but—unlike the BRIGHT 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, it found no short-term impact on attendance rates, math test scores, or French test scores (Dumitrescu et al. 2011). But an evaluation conducted seven years after the program was implemented found that the program raised enrollment by 10.3 percentage points and attendance by 13.6 percentage points (Bagby et al. 2017).[3]

[3] The IMAGINE program was later combined with a package of complementary interventions under the Niger Education and Community Strengthening (NECS) program, which were designed to increase access to high-quality education and improve reading achievement. As a result, the impacts estimated under the 10-year evaluation reflect the combined impacts of both the IMAGINE and NECS programs.

While relatively few studies have examined the impacts of infrastructure on the amount of time spent effectively on learning tasks during the school day, it is plausible that there is meaningful relationship between school infrastructure and effective instructional “time on task.” For example, if classrooms cannot maintain a comfortable temperature or adequate lighting, or if students are not able to access in-building sanitary facilities that are close to classrooms, these factors could reduce the amount of time that can be spent directly on instructional activities during the school day. There is also a recent and growing body of literature specifically linking measures of indoor and outdoor air quality to learning outcomes. For example, a large-scale, quasi-experimental analysis of changes in exposure to particulate matter (PM) in the United States found that 1-unit decrease in students’ exposure to $PM_{2.5}$ (that is, a decline of 1 microgram per cubic meter or air) produces an increase of 0.02 standard deviations in student test scores (Gilraine and Zheng 2022). Similar results have been reported in a European setting as well (for example, Palacios et al. 2022). Outside of air quality, it is less clear whether other school building improvements consistently 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 does increase learning time, evidence suggests that, in turn, this could produce important learning gains (as discussed in the literature review for the evaluation’s interim report; see Nichols-Barrer et al. 2019).

Importantly, we did not find any prior studies that assessed the impacts of school infrastructure in Georgia with rigor. 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 increasing the amount of time spent on learning activities can positively affect student learning, it is unclear whether in the Georgian context teachers will be able to use additional instruction time effectively to raise students’ test scores. This evaluation represents an important opportunity to fill these gaps in the research literature.

2. Prior evidence on training teachers and school directors

For the teacher and school director training activity, MCC’s cost-benefit analysis projected that this intervention would produce a 0.18 standard deviation improvement in student learning (in the medium-term and particularly in mathematics), ultimately resulting in a 2-percent improvement in students’ future annual earnings from employment (in the long term). Prior studies have shown that training interventions can have a wide range of potential effects. In addition, many of the strongest studies were carried out in contexts that are not directly relevant to the TEE activity, and it is not clear whether the effects seen elsewhere will be realized by the Compact. An overview of the relevant literature follows.

Prior studies have shown an uncertain relationship between training inputs for teachers and school directors and the outcomes targeted by the intervention. Some studies show strong effects, but others do not. In the United States, an extensive literature provides rigorous evidence demonstrating that variation in teacher quality is causally linked to improvements in students’ learning outcomes (for example, Chetty et al. 2011; Hanushek 2010). Rigorous studies of teacher training interventions in the United States also demonstrate that these interventions can have large effects on students’ learning in some circumstances (although evidence of impacts varies across programs). The evidence for these successful programs is concentrated in earlier grade levels, and the largest learning gains (in some cases, larger than 0.50 standard deviations) tend to be in studies of elementary school students in which the measured learning outcome aligned specifically with training materials (Yoon et al. 2007).

Evidence also exists from studies in developing countries that teacher training interventions can improve students' learning. Evans and Popova (2015) analyzed findings from six evidence reviews focused on education programs in developing countries. These evidence reviews summarized results from a total of 226 separate studies. The authors found suggestive evidence that extended teacher training programs that focus on pedagogical methods or academic subjects can have positive impacts on students' learning. In particular, the authors reported that longer-term trainings with ongoing follow-up support for teachers (the type of approach used in the one-year TEE training sequence) tended to outperform shorter-term (or one-time) training interventions with no follow-up mentoring or support. One example is the Read, Educate and Develop program in rural South Africa (Sailors 2010), which provided intensive professional development training for teachers, complete with demonstration lessons by mentors, monthly coaching visits by program staff, reflection sessions after monitoring visits, and after-school workshops for teachers. The study reported that the activity produced an improvement of 0.16 standard deviations in reading test scores. More broadly, the Evans and Popova (2015) review found that successful training and professional development interventions for teachers have had impacts on students' learning that range from 0.12 to 0.25 standard deviations.

Although we do not observe student-level outcomes in this evaluation of the TEE activity (due to the evaluation ending two years after the training was complete, which the program's designers believe would be too soon to expect changes in learning outcomes), we believe the literature provides a useful guide regarding the range of plausible effects that the program could initially produce on teachers' and school directors' practices. In our view, it is reasonable to assume that a change of a given size in students' learning would require at least a similar (if not substantially larger) change in measures of proximate teacher-level practices related to classroom instruction and pedagogy.

The existing literature examining the effects of teacher training programs, however, might not apply directly to the TEE activity on several counts. First, there have been no large-scale, rigorous evaluations of secondary-level teacher training programs in Georgia or other countries in post-Soviet settings. Most prior literature focuses on studies implemented in other regions, such as sub-Saharan Africa, Asia, and Latin America, where the teacher workforce likely differs substantially from that of Georgia with respect to formal education levels and pedagogical methods. Second, the focus of TEE is on education outcomes for students in grades 7 to 12, whereas most prior studies, including all 226 studies reviewed by Evans and Popova (2015), examine training of primary-level teachers. Substantial evidence suggests that learning outcomes are more difficult to affect in later grades relative to early grades (for example, see Hill et al. 2008), so the impacts found in early-grade interventions might not apply to TEE. Finally, the large-scale national rollout of the TEE activity makes it quite different from the smaller teacher training interventions that tended to be the focus of prior impact studies. Nearly all rigorous studies on this subject focus on small, targeted programs; for example, the average number of teachers trained in the evaluations reviewed by Popova et al. (2016) was 609. The current evaluation assesses a nationwide program that aimed to train up to 18,000 Georgian-language teachers and 2,085 school directors. Carrying out the TEE activity at such a scale could pose implementation challenges that were not present in the small interventions that have been the subject of evaluation studies in the past.

Another issue that may differentiate the TEE activity is the potential timeline for observing changes in teaching practice. During the planning process for this Activity, stakeholders and implementers designed the intervention with an understanding that the potential timeline for observing changes in teaching practice could be relatively slow (1–2 years at minimum). Implementers hypothesized that teachers would be unlikely to incorporate new lesson plans and teaching approaches immediately; instead, they would begin by piloting new lesson planning approaches during the first year and then enact changes more consistently over time once they found out which approaches appeared to be effective. We designed the evaluation and its data collection schedule to test the hypothesis that changes in teaching practices could occur over multiple years.

C. Objectives of the final report

This final evaluation report has two main objectives. For the evaluation of the school rehabilitation activity, which implemented a randomized controlled trial research design, the report presents a comprehensive set of impact estimates on outcomes including the quality of school infrastructure, the quality of the learning environment in classrooms and science labs, and students' academic achievement and attainment. These quantitative results are completed with rich qualitative data collected through a series of interviews and focus groups with school directors, teachers, and students exploring which aspects of school rehabilitation were the most important drivers of changes in learning. For the evaluation of the TEE activity, which implemented a descriptive analysis of trainees' instructional practices, the final evaluation builds on results from the 2019 interim evaluation report (Nichols-Barrer et al. 2019) and presents findings from a longer-term follow-up survey measuring the instructional practices of trained teachers 1–2 years after they completed the training sequence.

The remainder of this report is structured as follows. In Chapter II, we list the evaluation questions we seek to address, present the evaluation design, and describe the evaluation's data sources and analysis approach. The following chapters present our findings related to the school rehabilitation activity (Chapter III) and the TEE activity (Chapter IV). We conclude in Chapter V with a discussion of how our findings answer the evaluation's original evaluation questions and the implications of the findings for education policymakers in Georgia and in other country settings.

II. Evaluation Design and Final Analysis Approach

In this chapter, we review the design for the evaluation of the IGEQ Project and describe the analyses performed for this final report. We begin by listing the key evaluation questions and explaining how we used the evaluation to answer them. We then describe the data analyzed in the report—which include both quantitative and qualitative data—and our analysis approach.

A. Impact evaluation design for the school rehabilitation activity

Our evaluation for the school rehabilitation activity uses a mixed-methods study design with three components: (1) a process evaluation examining the program's implementation and costs, (2) an impact evaluation using a random assignment design to estimate the causal impacts of rehabilitation compared with a control group, and (3) a qualitative analysis of the relationship between changes in school infrastructure and changes in the learning environment in a subset of study schools. The evaluation's interim report presented results for the process evaluation (Nichols-Barrer et al. 2019); this final report presents results for the evaluation's impact study and qualitative analysis, drawing on data collected in annual follow-up data collection rounds carried out between 2017 and 2022.

Table II.1 presents the evaluation's key research questions. Our process evaluation (presented in the evaluation's interim report; see Nichols-Barrer et al. 2019) examined outcomes related to program design and implementation; the impact evaluation (the focus of this final evaluation report) examines the program's effects on school infrastructure, teachers, and students; and the study's in-depth qualitative analyses (also presented in this report) examine the relationships between rehabilitation inputs and the pattern of impacts observed in the quantitative study. The table also summarizes the data sources we used for each component of the research.

Table II.1. Final evaluation questions for the ILEI activity and approaches to answering them

Key evaluation questions	Evaluation components
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 students' and teachers' perceptions of health and safety? What are the impacts of the ILEI activity on teachers' behavior, such as attendance and time spent teaching?	• 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, in both treatment and control schools • 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 students' outcomes? What are the impacts on attendance, enrollment, dropout and retention rates, time spent studying in and out of school, and learning outcomes?	• Analyze teacher and student attendance, time on task, and teaching practices through survey 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)?	• Analyze administrative data on student attainment rates • Analyze if impacts of the ILEI activity are sustained three to five years after rehabilitation is complete

ILEI = Improved Learning Environment Infrastructure; RCT = randomized controlled trial.

1. Impact evaluation applying a randomized controlled trial (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 at minimum received rehabilitation design assessments, and—when rehabilitation was feasible—treatment schools received the program's full set of infrastructure rehabilitation services. As part of the Compact, GoG stakeholders agreed that schools assigned to the control group would only receive business-as-usual maintenance and operations support during the life of the five-year Compact (until July 2019). At the conclusion of the Georgia II Compact in July 2019, a total of 91 schools had been rehabilitated under the ILEI activity: 88 of the 91 rehabilitated schools are in the evaluation sample, and there were 3 rehabilitated schools that were not eligible for random assignment and were not included in the impact evaluation sample.[4]

To develop the random assignment procedure, we first stratified the sample of schools by region. Within regions that had enough schools, we further stratified the sample on the following school-level characteristics:

- 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 considered the design status of sampled schools in September 2014, when the first phase of random assignment took place. During the 2013–2014 school year, MCA-G hired a design contractor (Louis Berger) and completed rehabilitation designs for several schools in the Phase I regions. 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 had existing rehabilitation designs in September 2014. To realize cost savings from this prior design work, at the request of MCA-G and MCC, the evaluation gave the predesigned schools a higher probability of being assigned to treatment (66 percent) than the schools lacking pre-existing designs. To do so, our approach placed the pool of predesigned schools in its own separate set of region-level random assignment blocks. The study's impact analyses adjusts statistically for differences in the probability of selection into treatment associated with these predesigned strata.

This random assignment process took place in three phases that correspond to the program's staggered implementation schedule. Each of Georgia's regions was assigned to a different implementation phase (Table II.2)—this enabled the rehabilitation work in each phase to take place in a set of proximate regions, facilitating program logistics. At the beginning of a given phase, Mathematica randomly selected which schools were eligible to receive the program from a list of schools in each region that was vetted by MCC, MCA-G, and GoG stakeholders. This vetting process was primarily focused on the likely cost and feasibility of including a site in the program: for example, if a school needed major excavation work to repair problems with the building's foundation (suggesting the building might be a candidate for a complete tear-down and rebuild intervention, rather than rehabilitation), it was excluded from random assignment. Mathematica completed the random assignment process for schools in the Phase I regions in September 2014, for schools in the Phase II regions in July 2015, and for schools in the Phase III regions in July 2016. We collected baseline data in the first school year following randomization for schools in

[4] MCC's STAR report, available at https://www.mcc.gov/resources/pub-full/star-report-georgia-ii, provides more information on the full set of 91 rehabilitated schools.

each phase: 2014–2015 for Phase I schools, 2015–2016 for Phase II schools, and 2016–2017 for Phase III schools. As discussed below (Section 3), the timing of follow-up data collection activities in each school was aligned with the year rehabilitation work was completed.

Table II.2. Regional rollout of the ILEI activity

Phase	Regions	Number of treatment schools	Rehabilitation year
I	Mtskheta-Mtianeti, Racha-Lechkhumi and Kvemo Svaneti, Samtskhe-Javakheti, Shida Kartli	37	By December 2017
II	Kakheti, Kvemo Kartli	31	By December 2018
III	Guria, Imereti, Samegrelo-Zemo Svaneti	27	By December 2019

Note: In a few cases rehabilitation delays extended beyond the dates shown in this table. One phase 1 school was rehabilitated in 2018 and three phase II schools were rehabilitated in 2019.

ILEI = Improved Learning Environment Infrastructure.

2. In-depth qualitative research on the effects of school rehabilitation

In addition to the process and quantitative impact evaluation, our approach also includes 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 extensive, open-ended interview and focus group data. The qualitative analysis collected data in the second follow-up year after rehabilitation in each treatment school. In total, Mathematica selected a subset of about 10 percent of the schools in the impact evaluation sample (20 schools—10 treatment and 10 control), in which to collect in-depth, qualitative data about program implementation and results at these schools. The data collection paid particular attention to maintenance and operations practices, perceptions of school quality and safety, instructional time (referred to throughout this report as "time on task"), and the use of various school facilities. We acquired this information by conducting in-depth interviews with school directors and teachers and by discussing it with secondary school students in focus groups. The in-depth interviews with school directors assessed infrastructure usage patterns, school operations, and maintenance practices; the in-depth interviews with teachers assessed how school facilities are used, time on task, perceptions of school building quality and safety, and teacher attendance. The focus group discussions with students likewise assessed how school facilities are used, time on task, perceptions of school quality and safety, and determinants of student attendance.

These qualitative research activities are 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 are 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 focused 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 is 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.

3. ILEI study sample and data collection time frame

To align data collection with the key outcomes envisaged in the ILEI activity's program logic, we targeted data collection efforts to students who will be in grades 9 to 12 during the study's follow-up period. Specifically, in each school, we defined the baseline study sample to be all students enrolled in grades 8 and 10 in the baseline school year. We originally planned to reinterview the students in the baseline sample in later follow-up rounds. But because of implementation delays and uncertainty regarding the final school rehabilitation schedule, many of the grade 10 students interviewed at baseline would have aged out of secondary school by the time rehabilitation was completed. As a result, we abandoned the original longitudinal design and instead interviewed a new panel of students in the study's follow-up survey rounds. (We used the baseline data to calculate cross-sectional school-level covariates for the impact analysis; the study also used administrative data to track longitudinal patterns of enrollment and grade promotion across all grades.) The first follow-up data collection round surveyed all students enrolled in grades 9 and 11 in the year rehabilitation work was completed, and subsequent follow-up rounds surveyed all students in grades 10 and 12.

An MCA-procured local data collector (the Institute for Polling and Marketing, IPM) collected survey data, direct observations of attendance, and ratings of school infrastructure. The National Assessment and Examination Center (NAEC) developed and collected learning assessments for the study. Mathematica also obtained administrative data from Georgia's education management information system (EMIS) and implementation records.

For the evaluation's qualitative components, the study collected additional descriptive and qualitative data to investigate how rehabilitation affected the learning environment at study schools. For the qualitative data collection, we drew a sample designed to explore variation across a subset of schools in various regions throughout the country in the second follow-up year after rehabilitation work had been completed. For the analyses, we collected data from five regions that received the program towards the beginning of the Compact (the first regions where schools were rehabilitated in 2016 and 2017) and two regions that received the program towards the end of the Compact (where schools were rehabilitated in 2019). The qualitative sample comprised a balanced sample of treatment schools and control schools in each region, and we purposively selected schools to include a representative range of characteristics, such as school size and urbanicity. Within each of these schools, the local data collection firm conducted 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.

Table II.3 summarizes the data collection schedule. Because ILEI rehabilitation activities occurred in multiple phases, the data collection rounds occurred sequentially, by region (data collection for a given phase encompassed all treatment and comparison schools in the regions assigned to that phase). Note that because of program implementation delays, rehabilitation work in the Phase I regions originally scheduled to occur in summer 2015 was delayed until either 2016 or 2017. Similarly, rehabilitation work in Phase II schools originally scheduled for summer 2016 was delayed until 2018, and in Phase III schools, rehabilitation work originally scheduled for summer 2017 was delayed until the first half of 2019.

Table II.3. ILEI evaluation data collection schedule

Collection round (spring semester)	Phase I regions (rehabilitation completed in 2016)	Phase I regions (rehabilitation completed in 2017)	Phase II regions	Phase III regions
2015	Baseline data collection with grades 8 and 10 students	Baseline data collection with grades 8 and 10 students	None	None
2016	None	None	Baseline data collection with grades 8 and 10 students	None
2017	One-year follow-up with grades 9 and 11 students	None	None	Baseline data collection with grades 8 and 10 students
2018	Two-year follow-up with grades 10 and 12 students Qualitative data collection	One-year follow-up with grades 9 and 11 students	None	None
2019	None	Two-year follow-up with grades 10 and 12 students Qualitative data collection	One-year follow-up with grades 9 and 11 students	One-year follow-up with grades 9 and 11 students (schools completed by February 2019)
2020	None	None	Two-year follow-up with grades 10 and 12 students[1]	• One-year follow-up with grades 9 and 11 students (schools completed after February 2019)[1] • Two-year follow-up with grades 10 and 12 students (schools completed before February 2019)[1]
2022	Five-year follow-up with grades 10 and 12 students[2] Qualitative data collection	Four-year follow-up with grades 10 and 12 students[2]	Three-year follow-up with grades 10 and 12 students[2]	• Two-year follow-up with grades 9 and 11 students (schools completed after February 2019)[2] • Three-year follow-up with grades 10 and 12 students (schools completed before February 2019)[2] • Qualitative data collection

[1] 2020 data collection activities were delayed from spring 2020 to fall 2020 due to pandemic-related school closures.

[2] Due to school closures in 2020 and 2021, the count of "follow-up years" in this data collection round has been reduced by one year.

ILEI = Improved Learning Environment Infrastructure.

The data collection schedule in 2020–2022 was also affected by the COVID-19 pandemic in important ways. Schools in Georgia were entirely closed for most of March 2020, and they began requiring students to use remote (online) learning options from April 2020 until the end of the 2019-2020 school year. At the beginning of the 2020-2021 school year, about a third of school buildings remained closed and a larger proportion of schools enacted intermittent closures during the school year in response to local COVID-19 outbreaks. By the end of the spring 2021 semester, nearly all schools in Georgia were offering in-person learning again. In addition, for all the 2020-2021 and 2021-2022 school years social distancing restrictions were enforced in all of the schools that were open for in-person learning activities; these rules required students and staff to mask and remain several feet apart throughout the school day.

In the 2020 data collection round, due to school closures, survey activities planned for the spring were delayed until the subsequent fall. In addition, to account for the lack of access to school buildings during the pandemic, the evaluation also delayed the follow-up for the final cohort of Phase III schools from 2021 to 2022. To gather more data about the effects of the pandemic—alongside the longer-term impacts of improved school infrastructure—the 2022 data collection round was also expanded to include infrastructure assessments and learning assessments in all of the schools in the evaluation sample. This final data collection round constitutes a cross-sectional snapshot of these outcomes across all treatment and control schools, providing an opportunity for the evaluation to assess the longer-term effects of school rehabilitation (specifically, whether infrastructure upgrades had been adequately maintained and whether student learning outcomes changed over a longer period of time) three to five years after the upgrades had been completed.

A final contextual factor for this study is that the TEE activity was also rolling out a nationwide teacher training program concurrently with the school rehabilitation activity. For teachers in the treatment and control schools in the school rehabilitation study, the TEE training sequence ended either in September 2017, September 2018, or September 2019. Importantly, the TEE activity included all teachers in Georgia, regardless of whether they were in the school rehabilitation study's treatment group (working in a school assigned to receive rehabilitation) or in the control group (working in a school that was not assigned to receive rehabilitation). Because the impact estimates for the school rehabilitation activity are comparing results for TEE-trained teachers in treatment schools to the results for TEE-trained teachers in control schools, any effects of the TEE activity on teacher outcomes or student learning outcomes can be considered on a separate (orthogonal) basis, relative to the impact estimates for the school rehabilitation activity presented in this report.

4. ILEI study data collection and analysis approach

As part of the analysis, the ILEI evaluation collected baseline and follow-up survey data on the ILEI activity's key outcomes from students, parents, teachers, and school directors. The survey data were complemented by administrative data, study-administered learning assessments, and direct observations of student attendance and school infrastructure.

Mathematica developed five data collection instruments in English: survey questionnaires of students, parents, teachers, and school directors, as well as a school building infrastructure assessment. The full text of each of these data collection instruments can found in the evaluation's baseline report (Nichols-Barrer et al. 2017). The infrastructure assessment instrument provided the enumerators with consistent metrics for measuring school structures and systems. The infrastructure assessment teams were comprised of enumerators with engineering backgrounds who received training on how to consistently measure air quality, building systems, light levels, and temperature. Mathematica provided the technical measurement devices for this work and oversaw the training of the data collection team to ensure the protocols were carried out consistently. For example, Mathematica ensured that air quality inside classrooms was consistently measured in the same part of the classroom across all sites. Mathematica also oversaw that all air quality measurement devices, such as those for measuring levels of particulate matter and carbon monoxide, were used according to consistent protocols.

Analyses for this report use data collected across schools rehabilitated in each of the program's implementation years. Table II.4 summarizes the sample sizes for the treatment and control schools in each of these programmatic cohorts. To gather more data about the effects of the pandemic—alongside the longer-term impacts of improved school infrastructure—the 2022 data collection round was expanded to include infrastructure assessments and learning assessments in all of the schools in the evaluation sample (not just the schools rehabilitated in 2019, which were scheduled to receive two-year follow-up surveys, learning assessments, and infrastructure assessments in 2022). This final data collection round provides an opportunity for the evaluation to assess the longer-term effects of school rehabilitation (specifically, whether infrastructure upgrades had been adequately maintained and whether student learning outcomes changed over a longer period of time) three to five years after the upgrades had been completed.

To analyze year-to-year changes in student enrollment levels before and after rehabilitation, we also obtained administrative data from EMIS for all enrolled students at these schools. The data include anonymized information for each student who enrolled in one of the evaluation's treatment or control schools. The data include lists of enrolled students for nine school years, from 2013–2014 through 2021–2022 (the final school year in the evaluation period). The EMIS data also include records of each student's enrollment status in the subsequent school year (whether they remained enrolled in a rehabilitated school, dropped out of school, transferred to a school other than the rehabilitated school, or graduated from secondary school in grade 12). These measures of enrollment in the subsequent school year are available in all school years except 2021–2022.

Table II.4. ILEI follow up evaluation samples

Rehabilitation year	Survey round	Data collection dates	Survey data				
			Number of schools	Number of students	Number of teachers	Number of directors	Number of parents
2016	Year 1 follow-up	Feb 6 – Mar 17, 2017	20	844	147	20	795
	Year 2 follow-up	Feb 6 – Mar 17, 2018	20	798	144	20	747
	Year 5 follow up	Feb – Mar, 2022	20	732	n.a	n.a	n.a
2017	Year 1 follow-up	Feb 6 – Mar 17, 2018	38	1,991	313	39	1,851
	Year 2 follow-up	Feb 7 – Mar 27, 2019	38	1,836	294	37	1,742
	Year 4 follow-up	Feb – Mar, 2022	38	1,915	n.a	n.a	n.a

Rehabilitation year	Survey round	Data collection dates	Survey data				
			Number of schools	Number of students	Number of teachers	Number of directors	Number of parents
2018	Year 1 follow-up	Feb 7 – Mar 27, 2019	67	4,475	594	65	4,210
	Year 2 follow-up	Feb – Mar, 2020 / Sept – Oct, 2020	66	3,728	527	61	2,681
	Year 3 follow up	Feb – Mar, 2022	67	4,576	n.a	n.a	n.a
2019	Year 1 follow-up	Feb – Mar, 2020 / Sept – Oct, 2020	50	2,308	357	43	1,584
	Year 2 follow-up	Feb – Mar, 2022	51	2,766	431	45	2,291

Notes: For schools rehabilitated in 2016 and 2017, baseline data were collected between Apr 30 and Jun 7, 2015; for schools rehabilitated in 2018 baseline data were collected between Apr 1 and May 29, 2016; and for schools rehabilitated in 2019, baseline data were collected between Feb 6 and Apr 17, 2017.

ILEI = Improved Learning Environment Infrastructure.

The qualitative data used for the interim report came from 15 treatment and 10 control schools (25 schools in total). In each school, IPM's data collection team conducted one school director interview, four teacher interviews (grades 10 and 12; two science, one math, one foreign language), and two student focus groups (grades 10 and 12). Student focus groups included between 8 and 10 randomly selected students (with a random selection procedure designed to invite an equal number of boys and girls to participate). Data were collected in all 25 schools. Only 24 school director interviews were completed, however, because one director refused to participate in the study (Table II.5).

To collect these data, Mathematica's research team (1) trained interviewers and focus group moderators on best practices in qualitative data collection, (2) provided relevant background on the study goals, (3) explained in detail the respondent-specific qualitative instruments, and (4) oversaw practice sessions (role play and in the field). The four-day training for the qualitative data collection for field staff included a review of the background and purpose of the ILEI study, a detailed presentation of each qualitative protocol, role play and peer practice, and on-site practice sessions. Interview and focus group field practice took place in schools not in the study sample.

Table II.5. ILEI qualitative data collection sample

Qualitative data collection method	Respondent	Number of cases per school[a]	Total
In-depth interviews	School directors	1	24
In-depth interviews	Teachers (grades 10 and 12; 2 science, 1 math, 1 foreign language)	4	100
Focus groups	Students (grades 10 and 12)	2	50

[a] The sample included 25 schools in total: 10 schools that completed rehabilitation work in 2016, 5 schools that completed rehabilitation work in 2019, and 10 control-group schools that were not rehabilitated.

ILEI = Improved Learning Environment Infrastructure.

Before the school visit, IPM contacted schools to introduce the data collection activities and schedule interviews and focus groups. Interviews and focus groups took place at the schools and were audio recorded, transcribed verbatim, and translated into English. A small sample of the transcripts was randomly selected for quality assurance. Mathematica's Georgian consultant verified those translated transcripts against the audio recordings to check accuracy of the transcription and translation process. After completing these quality assurance reviews, Mathematica staff reviewed translated transcripts and imported approved transcripts into NVIVO (a qualitative analysis software package) for analysis.

a. Quantitative analysis

The impact analysis for the school rehabilitation activity relies on a random assignment design and compares the outcomes in rehabilitated schools to a control group of schools that was not rehabilitated. Due to feasibility and cost constraints realized after random assignment, seven of the 95 schools in the treatment group were not rehabilitated. To account for this, the analysis estimated intent-to-treat (ITT) impacts as well as treatment-on-the-treated (TOT) impacts for the school rehabilitation activity. While ITT estimates represent the average effect of being assigned to receive rehabilitation, TOT estimates represent the average effect of school rehabilitation on rehabilitated schools.

With the ITT approach, all schools that were randomly assigned to the treatment and control groups are included in the impact analysis and are analyzed in the groups to which they were assigned. We used an ordinary least squares (OLS) regression for this analysis (results are presented in Appendix A). For the TOT model, which provides the primary impact estimates for the evaluation, we used an instrumental variable approach where random assignment to the treatment or control group (95 schools were assigned to treatment whereas 81 were assigned to control) is used as an instrument for receiving rehabilitation (88 treatment schools were rehabilitated). To estimate TOT impacts, we use the following two-stage least squares (2SLS) regression models:

$$(1)\ \hat{R}_s = \beta_0 + \beta_1 TREAT_s + \beta_2 X_{is} + \eta_{ist}$$

$$(2)\ Y_{ist} = \alpha + \delta_1 \hat{R}_s + \delta_2 X_{is} + \varepsilon_{ist}$$

In the first stage, R_s is the rehabilitation dummy variable indicating whether a school was rehabilitated; $TREAT_s$ is the treatment dummy variable indicating whether a school was randomly assigned to receive treatment; X_{is} includes a set of student-level demographic characteristics, an initial difference in school infrastructure found at baseline, and a set of school-level variables defining the random assignment blocks; and finally, η_{ist} is the random error. This model estimates the effect of treatment status on school rehabilitation, while controlling for X_{is}. The predicted value of $\hat{R}_s$ is used for the second stage model. In the second stage, 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. (Impacts were estimated separately for each outcome year; we focused our analysis on the second follow-up year, but impacts for the first follow-up year are presented in Appendix B.) X_{is} is the same set of covariates used in the first stage, and finally, ε_{ist} is the random error. The estimated value of the coefficient δ_1 represents the impact of rehabilitation on the outcome of interest (under the assumption that random assignment to the treatment group had no relevant effects on the treatment schools that were not rehabilitated). Standard errors in the model were clustered at the school-level using the standard Huber-White estimator to account for the possibility of correlations among individuals' characteristics within schools.

Subgroup analyses. We also conducted several subgroups analyses. For example, we explored whether the impacts of school rehabilitation differ for girls versus boys, or for the subgroup of schools where we collected outcome data before the COVID-19 pandemic versus during the pandemic. To do so, we used the same 2SLS regression model above, but with a second stage that included a binary indicator for the subgroup of interest and an interaction term between the treatment indicator and the subgroup indicator. In this model, the coefficient on the interaction term represents the additional effect of being in a particular subgroup on the outcome interest (for example, the treatment effect for girls), relative to a reference group (the treatment effect for boys).

Infrastructure outcome indices. As we did for the analyses conducted for the baseline and interim reports, we constructed indices for most aspects of school infrastructure measured in the final surveys. Data reduction was necessary to avoid the multiple comparisons problem, which arises when researchers report the results of many hypothesis tests, some of which are bound to be falsely rejected due to pure chance. This is the same logic whereby flipping a coin many times will eventually yield streaks of all heads or all tails, even if the coin flip is fair.

To define the key outcome indices for the evaluation, we used principal components analysis (PCA) to combine multiple measures related to aspects of school infrastructure into a single index.[5] Each index is a weighted average of related infrastructure measures, in which the weights are aligned with measures with the highest component scores (that is, an infrastructure measure that explains a greater amount of variation across schools will receive a larger weight than measures explaining less of the variation in the sample). We further standardized the indices within the sample of schools to *z*-scores, so each index has a mean of 0 and a standard deviation of 1. Although the specific values of the indices cannot be directly interpreted, each index was coded to represent the presence of infrastructure gaps or problems and can be used to compare the infrastructure in treatment and control schools. For example, a school with a higher score on the index of physical classroom conditions would have worse conditions than a school with a lower score.

To maintain comparability to the baseline results, we used the PCA weights estimated at baseline to construct the final indices and used the maximum values of each variable at baseline to standardize the final indices. For more details on the weights for each index included in the baseline, interim, and final index construction, please consult the baseline report (Nichols-Barrer et al. 2017). The constructed indices used in the impact study are below:

- *Better condition of school building exterior***.** Includes measures of the condition of the school building roof, the condition of the rainwater drainage system, the condition of main entrance doors, and whether the exterior of the building is painted.
- *Better condition of interior structures***.** Includes summary measures of the condition of the walls, ceilings, and floors in all classrooms and the indoor gym (if present).

[5] A PCA is a statistical procedure that determines how a number of factors (in our case, related measures of infrastructure) are correlated with one another and condenses this information into linear combinations of the factors, called principal components. Each principal component consists of a number of weights, or factor loadings, that define how much of the variation in the principal component is driven by each factor. We adopted the weights estimated for the first principal component to calculate our indices because, by design, the first principal component contains the set of factor weights that captures as much of the correlation between the factors as possible.

- *Better condition of stairs in main school building***.** Includes measures of the condition of the stairwells in the main school building, whether the stairs are level, and whether the stairs are evenly spaced (if two or more floors are present in the main school building).
- *Better air quality in classrooms***.** Includes measures of the presence of particulate matter equal to or smaller than 2.5 microns in width ($PM_{2.5}$) and between 2.5 and 10 microns in width (PM_{10}) in parts per million (ppm) (there is extensive evidence that exposure to particulate matter can have negative health consequences; World Health Organization 2013), the presence of carbon monoxide in ppm, and whether smoke was visible in the classroom.
- *Better condition of classroom teaching facilities***.** Includes measures of whether all classrooms in a school have working lights, a lockable door, and a blackboard visible from the back of the classroom, as well as the condition of teaching equipment in classrooms.

b. Qualitative analysis

The research team developed a coding scheme to identify and parse meaningful segments of transcripts linked to the key qualitative research questions and inventory-related themes and findings across respondents. After all qualitative data were coded, the research team exported data by code and systematically reviewed the qualitative evidence pertaining to the study's research questions. Analysis focused on identifying consistent patterns and trends across transcripts (by respondent and across respondents). We also identified outliers or respondent disagreements in relation to a key theme or pattern. We documented analyses and triangulation of findings across respondents in memos and summary tables with illustrative quotes. Appendix C contains a master table that summarizes qualitative findings for the school rehabilitation study, with illustrative quotes.

B. Evaluation design for the TEE activity

For the TEE evaluation, our evaluation design relies on quantitative surveys and qualitative data collection methods to examine the potential effects of the training initiative.

The evaluation's interim report (Nichols-Barrer et al. 2019) presented results from a mixed-methods study design that included two components: (1) a performance evaluation that assessed the possible effects of the TEE activity on school management and classroom instructional practices using descriptive surveys and qualitative data and (2) a matched comparison group design that assessed the initial impacts of the Activity's teacher training modules, also using survey data. The performance evaluation and the matched comparison group analysis were designed to answer research questions about the program's implementation and initial outcomes; we used evidence from these analyses to assess whether the program had plausible effects on teachers' and school directors' practices that could, in turn, produce gains in students' learning and longer-term labor market outcomes.

This final evaluation report extends the analyses presented in the evaluation's interim report through a descriptive analysis of trends in teachers' attitudes towards (and use of) a wide range of enhanced instructional practices. Specifically, because the TEE training sequence took place on a nationwide basis, the evaluation's matched comparison group design could only examine the potential effects of the program during a narrow window of time (shortly after the first training cohort had completed the year-long training sequence, and shortly before the second training cohort began its own training sequence). This snapshot of teacher practices may not have allowed enough time for teachers to test out and apply the information provided through the program. To address this issue, the final analysis collected an additional round of survey data from both cohorts of teachers, 12 months later. For the first training cohort, this represented a two-year follow-up after training was completed. For the second training cohort, this represented a one-year follow-up after training was completed.

Table II.6 presents the research questions that each component of the TEE evaluation investigated. The specific research questions that we examine in this final report are highlighted in bold.

The performance evaluation collected information about how the TEE activity was implemented, tested whether program activities were implemented as designed, and assessed whether the practices of trained teachers and school directors align with the activities' targeted set of practices related to classroom instruction and school management. The performance evaluation analyzed several different types of data using multiple data sources, including program documentation, survey data, and qualitative research.

As part of the interim evaluation report, the study used project reports and training databases to document the set of activities delivered (for example, the number of teachers and school directors trained and the number of schools receiving ongoing support from members of the project's training teams). To understand how the program might affect training participants and how they apply new information and skills to their work, we also collected survey data from a representative sample of teachers and school directors trained by the program. Ultimately, we collected survey data at three points in time: September 2017 (one to four weeks after the first cohort of teachers completed its sequence of four training modules), September 2018 (one to four weeks after the second cohort completed its full sequence of TEE trainings), and a September 2019 survey round focused on teachers' longer-term outcomes (two years after the first cohort of teachers completed its training sequence and one year after the second cohort completed the training program).

The performance evaluation also used qualitative data to understand how the program was implemented and how the program might have changed participants' practices. This included observation and monitoring of teacher study groups during the program's first implementation year (the 2016–2017 school year) to measure the extent of teacher participation in study group meetings. In addition, the evaluation used exploratory, in-depth interviews with school directors and focus groups with teachers during the program's second year (after the first cohort of teachers and school directors completed the Activity's full course of training modules) to gather more information about how the training was implemented and identify possible relationships between training activities and potential outcomes. Finally, the study directly observed the classrooms of a sample of trained teachers delivering lessons during a regular school day. These observations occurred during the program's second implementation year (the 2017–2018 school year) and focused on a small sample of 22 teachers who completed the training sequence in September 2017 and also participated in the evaluation's teacher survey. Results from these qualitative analyses were the focus of the evaluation's interim report.

Table II.6. Evaluation questions for the TEE activity and approaches to answering them

Evaluation questions	Approaches to answering them
Describe program design and implementation	**Performance evaluation**
Did the training activities embody a clearly developed theory of change? Did the TEE activity align with improvement goals and target pedagogical weaknesses identified by earlier research?	• Review program design documents, training materials, and implementation records

Evaluation questions	Approaches to answering them
Was the Activity implemented as designed? What were the main challenges to implementation? Was the amount of training uniform across cohorts and subject areas? What activities did school-based professional development facilitators undertake? Did teacher study group activities occur as designed?	• Use implementers' data to compare planned time lines, budgets, and work plans to actual activities • Conduct in-depth interviews with implementers and school-based professional development facilitators
Describe teacher and school director outcomes	**Performance evaluation**
To what extent do school directors perceive that their instructional leadership and school management skills have changed as a result of the new training interventions, including project-supported collaboration with other directors in their region? Do directors report changes in attitudes toward parental engagement and community engagement? **To what extent do teachers perceive that their pedagogical and classroom management practices have changed as a result of the new training interventions, project-supported collaboration with other teachers, and professional support from SPDFs?**	• **Analyze survey data collected from trained teachers** • Analyze survey data collected from trained school directors • Analyze survey data collected from students of trained teachers • Conduct focus groups with teachers and in-depth interviews with school directors to understand perceptions of changes in performance and behavior • Analyze classroom observation data for a subsample of trained teachers to describe pedagogical practices • Triangulate observational data on teachers' practices with self-reported teacher survey data and student survey data
To what extent have school directors' instructional leadership and school management practices improved? **To what extent have teachers' pedagogical practices (for example, using student-centered instruction, matching practice to subject matter, using formative assessment) and classroom management (for example, using affirmative teaching, eliminating gender bias, increasing time on task) improved?** To what extent do students experience student-centered instruction, formative assessment use, and classroom management practices that align with the goals of the teacher training activities (such as affirmative teaching, reducing gender bias, and engaging effectively with science facilities)?	
Assess potential effects of training on teachers	**Matched comparison group design**
Did teacher training modules improve teachers' knowledge about student-centered instruction, formative assessments, and classroom management? **Did teacher training modules improve teachers' willingness to use student-centered instruction, formative assessments, and classroom management?**	Compare the survey outcomes of teachers trained in 2016–2017 school year (Cohort 1) to a matched comparison group of teachers who will not be trained until the 2017–2018 school year (Cohort 2) **Track the outcomes of Cohort 1 and Cohort 2 teachers for 13–25 months after completion of training**

SPDF = School Professional Development Facilitators; TEE = Training Educators for Excellence. Research questions assessed in this final report (and the approaches used to answer them in this report) are highlighted in bold.

The performance evaluation did not include student learning assessments or student exams because the expected timing of any changes in student learning outcomes did not align with the timeframe of the evaluation's TEE-related surveys in 2017-2019 (and the Georgian government did not conduct universal examinations of secondary students in 2020-2022). As a result, the evaluation did not directly measure student learning outcomes. But because of concurrent data collection activities related to the evaluation of school rehabilitation activities, it was possible to collect descriptive data from students about their perceptions of teaching practices (using a convenience sample of students surveyed in spring 2018 as part of the school rehabilitation study). As part of the interim report, we used this survey data to measure students' perceptions related to teachers' use of student-centered instruction, formative assessments, and positive classroom management practices.

This final evaluation report builds on the results of the interim report by tracking teachers' attitudes and classroom instruction practices for an additional year after the training sequence ended. This analysis of trends is particularly important for assessing whether the program's theory of change is likely to be working as designed, with initial changes in teacher knowledge and attitudes towards improved teaching practices ultimately producing longer-term changes in classroom instruction over time.

1. Final analysis of trends in teacher outcomes

We used a descriptive evaluation design to examine longer-term changes in teachers' outcomes, 13–25 months after completion of the training sequence. Any effort to directly estimate program impacts involves comparing outcomes for a group of participants with outcomes for a comparison or control group that does not receive the same activity in a given time period. In the interim report, we endeavored to apply this type of design to evaluate the training program at a point in time when a group of teachers who were trained during the 2016–2017 school year (Cohort 1) could be compared with a group of teachers who were trained later, in the 2017–2018 school year (Cohort 2). This type of comparison group analysis was not feasible for the analysis of trends presented in this final report, because the data collection occurred after both cohorts of teachers had completed the training program.

That said, it is not clear that the matched comparison group findings discussed in the evaluation's interim report represent the most policy-relevant point in time to measure the effects of the training sequence. Changing classroom instructional practices takes time—teachers must learn how to pilot and refine new techniques and approaches, and it may require a full school year (or longer) before teachers can accurately assess whether a new set of practices is useful and whether they intend to continue using those practices in the future. To address this issue, this final evaluation report presents an analysis of trends in teachers' use of various instructional practices, including data up to two years after the training sequence ended. The trends analysis consists of simple descriptive comparisons of the average self-reported instructional practices used by teachers at two points in time: (1) immediately after the training sequence ended (fall 2017 for Cohort 1, fall 2018 for Cohort 2) and (2) in fall 2019 (25 months after training for Cohort 1, 13 months after training for Cohort 2). Because the follow-up period differs by cohort, the trend analysis was conducted separately for each cohort rather than pooling results across the full sample, or testing for statistically significant differences across cohorts.

2. TEE study population and evaluation sample

To identify teachers for the survey sample, the evaluation randomly selected a geographically representative sample of 120 schools, and in each school surveyed the teachers in upper grades (8 to 12) in the targeted subjects of English, geography, mathematics, and science. To reduce burden on study participants (specifically the time that would have been required to respond to survey questions related to the TEE activity alongside survey questions about the school rehabilitation activity), the sample frame also excluded all of the treatment and control schools participating in the school rehabilitation evaluation—there is no overlap between the TEE study sample and the school rehabilitation study sample. This has the added benefit of ensuring that the trends in outcomes we describe for teachers in the

TEE sample cannot be conflated with the potential impacts of the school rehabilitation activity (such as increases in time on task among teachers in rehabilitated schools).

These final analyses for the TEE evaluation focus on describing the outcomes of training activities delivered to two cohorts of teachers: the first two cohorts to receive training activities in Georgian-language schools during the 2016–2017 school year and the 2017–2018 school year. Although the TEE activity is nationwide in scope and ultimately included minority-language schools in later years, the two initial cohorts of trainees prioritized staff at Georgian-language schools. Thus, the study population is limited to all Georgian-language teachers in Georgian-language schools.

The TEE activity initially prioritized teachers who had passed a certification exam for their teaching subject (these teachers are classified as "senior," "lead," or "mentor" teachers), and the remaining openings in the first training cohort were offered to teachers who had not passed the certification exam for their subject (classified as "practitioner" teachers). Ultimately, 58 percent of Cohort 1 consisted of highly qualified senior, lead, or mentor teachers, and the remaining 42 percent of the cohort consisted of less qualified practitioner teachers. In contrast, nearly all of the participants in the second training cohort were practitioner teachers. In Georgia practitioner teachers (with an average age of 52) are older than teachers who have passed their certification exam (with an average age of 46). In other words, the first cohort of trainees was both younger and more likely to have a strong grasp of their teaching subject than teachers in the second cohort.

For the trend analyses in the final report, which do not rely on a matched comparison group design, the analysis sample included all of the trained teachers who competed the survey. Relative to the matched comparison group analyses in the interim report (which were limited to the subset of practitioner teachers in Cohort 1 who could be successfully matched to a comparable group of Cohort 2 practitioner teachers), this final analysis provides a more comprehensive picture of how the general population of all trained teachers in each cohort responded to the program.

3. TEE evaluation time frame

We collected survey data from the study's sample of school directors, Cohort 1 teachers, and Cohort 2 teachers at three points in time: late September 2017 (one month following completion of the first teacher cohort's training modules), late September 2018 (one month following completion of the full training sequence for Cohort 2), and late September 2019 (when we conducted a final data collection with Cohort 1 and Cohort 2 teachers to measure longer-term post-training outcomes). The Georgian academic year begins in September and ends the following June, meaning these surveys took place approximately one month after the school year began. Collecting data in September of each year provided an opportunity to assess teacher knowledge and attitudes towards TEE-related teaching practices shortly after the training sequence ended, and then assess how those outcomes changed one and two years later as teachers had an opportunity to apply the training in their classrooms over time. The evaluation also conducted qualitative data collection activities in a subsample of schools during the 2017–2018 school year to further investigate possible effects of the full training sequence on the first cohort of teachers and school directors (Table II.7).

An additional factor affecting the trend analyses in this report is that Cohort 1 teachers were given an opportunity to make up for any training modules that they missed and sit in on training events attended by Cohort 2 teachers. As a result, the percentage of Cohort 1 teachers who attended any TEE training modules increased from 64 to 82 percent after the second round of training. For a small portion of teachers, this additional exposure to training among Cohort 1 teachers may have affected their teaching practices in later years on a different schedule than the rest of the cohort. In other words, a subset of Cohort 1 teachers received at least some of their training sequence on a delayed schedule that more closely resembles the training schedule offered to Cohort 2.

Table II.7. TEE survey data collection schedule

Data collection round	Cohort 1 teachers	Cohort 2 teachers
September 2017	Initial outcome survey (*1 month after training)* N=877	Baseline survey (*before training)* N=309
September 2018	Year 1 outcome survey (*13 months after training*) N=784	Initial outcome survey (*1 month after training)* N=266
September 2019	Year 2 outcome survey (*25 months after training)* N=719	Year 1 outcome survey (*13 months after training*) N=220

TEE = Training Educators for Excellence.

III. Findings for the ILEI Activity

A. School rehabilitation program context

At the conclusion of the Georgia II Compact in July 2019, a total of 91 schools had been rehabilitated under the ILEI activity: 88 of the 91 rehabilitated schools are in the evaluation sample, and there are 3 rehabilitated schools that were not eligible for random assignment and were not included in the impact evaluation sample.[6] Among the schools in the evaluation sample, 12 were completed in 2016, 17 were completed in 2017, 34 were completed in 2018, and 25 were completed in 2019. Most of the rehabilitated schools (20 percent) were in Shida Kartli and Kakheti, 19 percent were in Imereti, 15 percent were in Kvemo Kartli, 7 percent were in Samegrelo-Zemo Svaneti, and 5 percent were in Samtskhe-Javakheti and Racha-Lechkhumi (Table III.1).

In accordance with the program's targeting criteria for eligible schools, prior to rehabilitation, all treatment schools had substantially higher enrollment and were more fully utilizing their school buildings compared to other schools in rural areas of Georgia. On average, rehabilitated schools enrolled a total of 405 students prior to rehabilitation (notably higher than the average of 159 students in Georgia's rural schools), and rehabilitated schools had an average of 5.9 square meters of indoor space per enrolled student (compared to 11.4 square meters per student in other rural schools). The proportion of socially vulnerable students in rehabilitated schools (25 percent) was similar to the percentage found in other rural schools in Georgia (30 percent).

Table III.1. Summary of baseline characteristics in rehabilitated schools

	Rehabilitated treatment schools	All schools in rural areas
Number of schools	88	1,567
Average total enrollment	405	159
Average school building size (m^2)	2,376	1,807
Ratio of school building size (m^2) to school enrollment	5.9	11.4
Percentage of socially vulnerable students	25%	30%
Average number of socially vulnerable students	80	48
Regional distribution of schools (percentage)		
Adjara	0%	12%
Guria	3%	5%
Imereti	19%	20%
Kakheti	20%	10%
Kvemo Kartli	15%	14%
Mtskheta-Mtianeti	2%	4%
Racha-Lechkhumi and Kvemo Svaneti	6%	3%

[6] MCC's STAR report, available at https://www.mcc.gov/resources/pub-full/star-report-georgia-ii, provides more information on the full set of 91 rehabilitated schools.

Table III.1 (*continued*)

	Rehabilitated treatment schools	All schools in rural areas
Samegrelo-Zemo Svaneti	8%	13%
Samtskhe-Javakheti	6%	12%
Shida Kartli	20%	7%

Note: Average total enrollment, average school building size, and percentage of socially vulnerable students were estimated using 2014 administrative education management information system data. The sample of other schools in rural areas of Georgia summarized in this table excludes schools in the cities of Tbilisi and Batumi (because urban areas are not eligible for the program) and schools in the disputed regions of Abkhazia and Tskhinvali. Schools outside Batumi in the Adjara region are excluded from the evaluation because implementers, Millennium Challenge Corporation, and Millennium Challenge Account-Georgia decided to exclude the region from random assignment.

B. Effects on school infrastructure

1. Physical condition of the school building

The first step in the theory of change of the ILEI activity is to improve the quality of school infrastructure. To assess building quality, the evaluation sent teams of two or three engineers to visit each school in the sample approximately two years after rehabilitation was completed, and then visit schools again in spring 2022 for the evaluation's endline data collection round. These teams carried out detailed visual inspections of the school building exterior, shared interior spaces (hallway, staircases, etc.), and classrooms to look for major infrastructure problems, such as visible gaps or holes, peeling paint, or unsafe stairways with uneven steps or missing handrails. They also assessed the building's heating, lighting, air quality, and sanitation systems using standardized, quantitative rubrics.

To summarize data from these infrastructure assessments, we created indices using infrastructure measures from the two-year follow-up building survey. Using a relatively small number of indices (rather than examining many indicators individually) reduces the risk of falsely attributing statistical significance to differences that actually result from chance alone. Each index summarizes the condition of a particular aspect of the school's infrastructure (for example, conditions in the school building's exterior) and can be used to measure differences between treatment and control schools in standard deviation units (commonly referred to as *z*-scores). Specifically, all the indices for school infrastructure are coded such that increasing index scores indicate better infrastructure conditions, and each index has been standardized to have a mean of 0 and a standard deviation of 1 for the schools in the evaluation sample. More details on the weights for each index can be found in the evaluation's baseline report (Nichols-Barrer et al. 2017).

Rehabilitation investments produced substantial improvements in the physical infrastructure of rehabilitated schools. Treatment schools experienced large and statistically significant improvements in the overall condition of the school building. For example, rehabilitation produced a 1.7 standard deviation improvement in the evaluation's index of wall, ceiling, and floor conditions (as measured in the school's classrooms and indoor gym). This is a large change. Standard deviation units represent differences in a school's percentile rank for infrastructure quality. In this case, a change of 1.7 standard deviations is equivalent to moving a school from the 21st percentile to the 79th percentile for infrastructure quality in the study sample. Treatment schools also experienced a major improvement in the condition of indoor stairs in the main school building, with a statistically significant impact of 1.4 standard deviations (equivalent to an increase from the 25th to 76th percentile). Treatment schools also experienced a smaller improvement in the condition of the school building exterior (an impact of 0.3 standard deviations, significant at the 10 percent level). As shown in Table III.2, we also observed that far more treatment than control schools have access to a working science lab (about 95 percent of treatment schools compared to 31 percent of control schools), but access to indoor gyms and outdoor recreation facilities was similar in the two groups.

Table III.2. Impact of rehabilitation on school infrastructure and teaching facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Better condition of school building exterior (*z*-score)	0.15 s.d	-0.15 s.d.	0.30^ s.d.	0.083	0.171
Better condition of walls, ceilings, and floors in all classrooms and indoor gym (*z*-score)	0.82 s.d.	-0.84 s.d.	1.67** s.d.	0.000	0.085
Better condition of stairs in main school building (*z*-score)	0.69 s.d.	-0.70 s.d.	1.39** s.d.	0.000	0.113
School has an indoor gym (p.p.)	84.5	78.9	5.5	0.349	0.059
School has an outdoor recreation area (p.p.)	64.5	72.7	-8.2	0.253	0.072
School has a science laboratory (p.p.)	95.3	31.2	64.1**	0.000	0.063

Source: Two-year follow-up building survey administered in 175 schools.

Notes: The first three rows show data on building condition indices with a mean of zero and a standard deviation of 1 (z-scores), with means and differences calculated in standard deviation (s.d.) units.

Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

p.p. = percentage point.

The program was highly effective in addressing problems with classroom walls, ceiling, and floors. By the two-year follow up, rehabilitation greatly reduced the percentage of schools that had a classroom with three or more problems with walls (-74 percentage points), ceilings (-73 percentage points), or floors (-58 percentage points), as shown in Table III.3. These types of problems were nearly eliminated in treatment schools—for example, none of the treatment schools had a classroom with three or more visible ceiling problems after rehabilitation (compared with 73 percent of control schools). The results are similarly strong for an alternative version of the outcome identifying classrooms with two or more problematic conditions in each of these areas, rather than three (results not shown).[7]

[7] Few rehabilitated compared to control schools had at least one classroom with two or more visible problems in ceilings (2 versus 91 percent), floors (5 versus 82 percent), and walls (38 versus 94 percent).

Table III.3. Impact of rehabilitation on problematic conditions in walls, ceiling, and floors of classrooms

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of schools that have at least one classroom with:					
Three or more problems in walls	6.3	79.8	-73.5**	0.000	0.055
Three or more problems in ceilings	0.0	72.6	-73.0**	0.000	0.055
Three or more problems in floors	2.5	60.7	-58.2**	0.000	0.057

Source: Two-year follow-up building survey administered in 175 schools.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

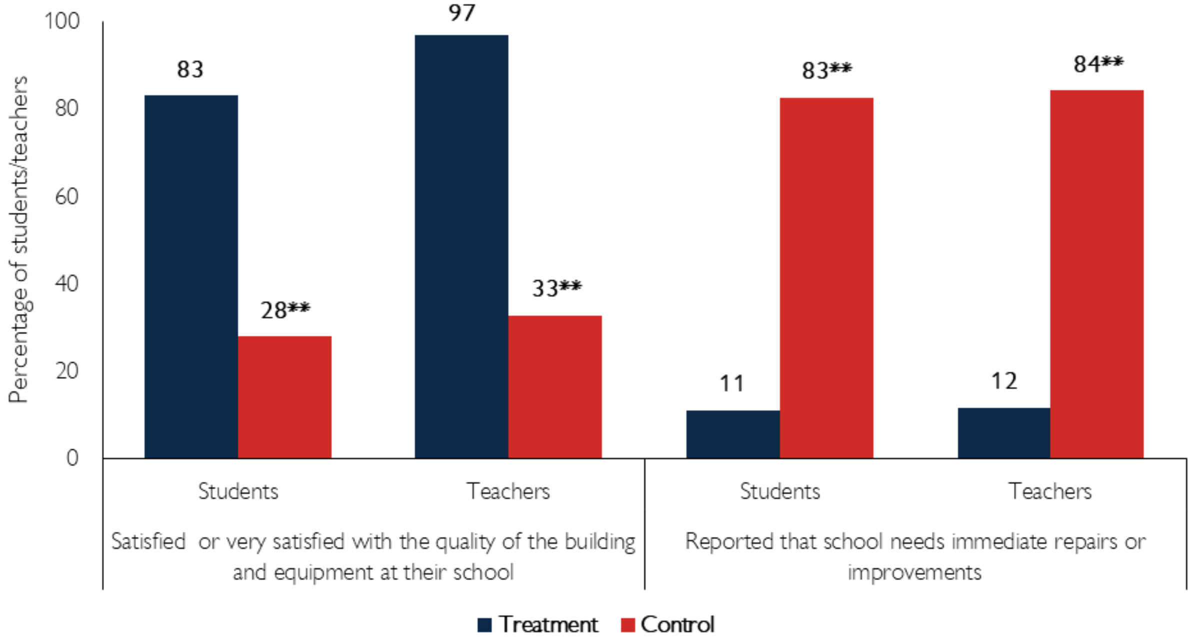
These improvements in school infrastructure were highly visible to students and teachers and produced large improvements in their levels of satisfaction with the building. Most of the teachers and students (97 and 83 percent, respectively) from rehabilitated schools reported that they are very satisfied with the quality of the building and equipment at their schools. In contrast, about one-third of teachers and students surveyed from control schools reported that they are very satisfied with the quality of the building and equipment at their school. Rehabilitation also produced a large improvement (of about 70 percentage points) on the percentage of teachers and students who reported that their school needed immediate repairs. Only 11–12 percent of teachers and students from treatment schools reported that their school needed immediate repairs, compared to 83–84 percent in the control group (Figure III.1).

Qualitative data confirm that addressing these types of highly visible infrastructure problems was important to students and teachers. This was particularly evident in control schools that were not rehabilitated. Almost all the teacher interviews and student focus groups in control schools revealed strongly negative views about school infrastructure quality. Respondents in these schools focused on the fact that their schools have damaged stairs, walls, roofs, floors, and windows. For example, students at one control school reported that they cannot close cracked windows and that there are visible holes in the building's walls and floors that expose them to drafts and outdoor weather. In contrast, respondents in rehabilitated schools consistently expressed pride in the quality of the school building, with some students noting that they were particularly happy to attend a school building that is of a higher quality than other schools in the surrounding region.

"We feel more comfortable obviously and do not have to worry about the cold now, the wind blowing through the broken glass. In a sense of the physical environment has improved quite a bit, not even improved, it has radically changed, and this will obviously be reflected on the learning process as well."

–Director, rehabilitated school in Guria◢

Figure III.1. Impact of rehabilitation on student/teacher perceptions related to the quality of physical building



Source: Teacher and student surveys completed by 1,380 teachers and 8,460 students, interviewed at two-year follow-up.

Notes: Inside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. Next to the bars are the differences between treatment and control mean. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

2. Effects on heating systems

Rehabilitation improved school heating systems and made classrooms more comfortable during winter months. Rehabilitation increased the availability of central heating systems by 36 percentage points, a statistically significant impact. Almost all treatment schools (98 percent) had functional central heating compared with 62 percent of control schools. As a result, treatment schools were much more likely to benefit from central heating in all classrooms (an impact of 31 percentage points) and indoor gyms (an impact of 36 percentage points). Importantly, rehabilitation also improved classroom temperatures in winter months. The evaluation team collected classroom temperature data in the month of February, to assess whether heating systems were being used effectively to keep classrooms warm during cold weather. On average, the median classroom in rehabilitated schools reached a warmer temperature (19°C, equivalent to 66°F) than the median classroom in the control group (16°C, equivalent to 60°F). The impact of rehabilitation on classroom temperatures (3.1°C, 6°F) was statistically significant as well (Table III.4).

Table III.4. Impact of rehabilitation on presence and perception of central heating

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Have functional central heating	97.9	62.3	35.6**	0.000	0.057
All classrooms have functional central heating	93.4	62.3	31.1**	0.000	0.061
Indoor gym has central heating	100.0	64.3	35.8**	0.000	0.064
Average measured temperature (median classroom, degrees in Celsius)	19.1	16.0	3.1**	0.000	0.487

Source: Building survey administered in 175 schools at two-year follow-up.

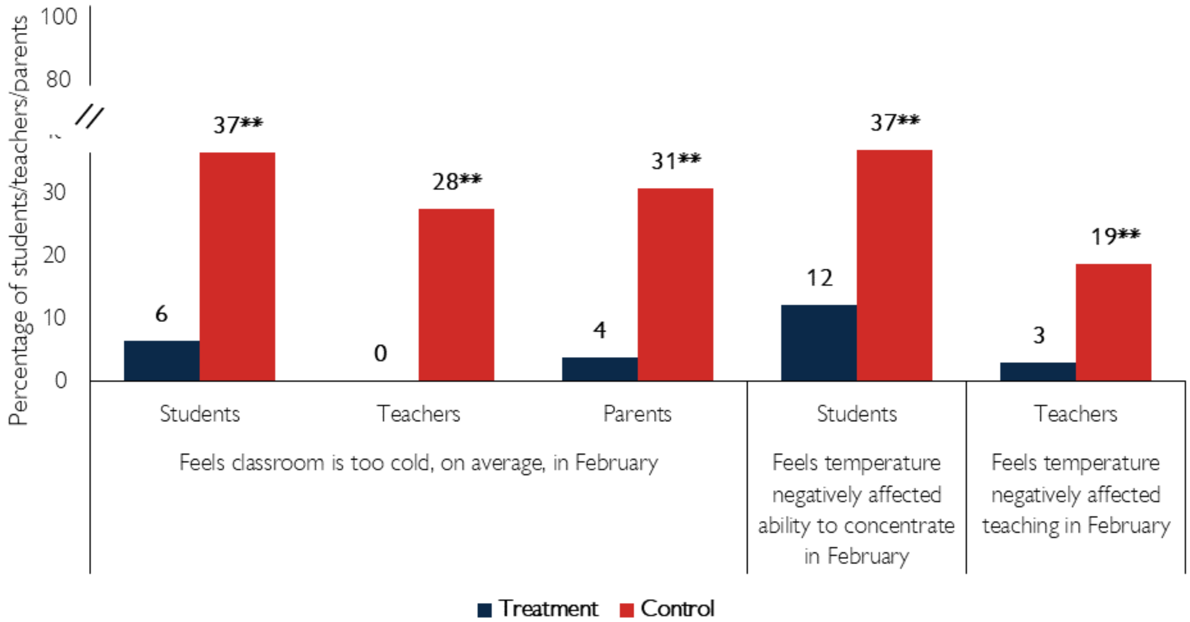
Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Students, teachers, and parents felt that central heating systems improved the learning environment in school. Introducing central heating greatly reduced the percentage of students, teachers, and parents who felt that classrooms were too cold in February. In control schools, about 37 percent of students, 28 percent of teachers, and 31 percent of parents reported that their classroom was too cold, whereas in rehabilitated schools, the percentage of respondents reporting these problems dropped to 0–6 percent (Figure III.2). Moreover, students and teachers both reported that improved temperatures improved the learning environment in winter. Rehabilitation substantially reduced the percentage of respondents who reported that classroom temperatures negatively affected teach and learning, with an impact of -25 percentage points for students and -16 percentage points for teachers (both statistically significant). In control schools, about four in 10 students reported that cold temperatures affected their ability to learn in winter, compared with only one in 10 students in rehabilitated schools. Qualitative data affirmed that addressing these heating issues was a high priority for students and teachers. In control schools, many respondents reported that, due to cold classrooms, it was necessary to wear thick coats during lessons, and in some cases, students and teachers resorted to standing up during class, huddling for warmth, or buying personal heating devices. None of these issues remained a problem in rehabilitated schools: in fact, one director explained that after rehabilitation, students and teachers appeared much more interested in staying in the building beyond the end of the school day, because they do not have access to such comfortable heating at home.

Figure III.2. Impact of rehabilitation on student/teacher/parent-perceived cold and its effect on learning environment



Source: Teacher and student surveys completed by 1,380 teachers and 8,460 students, interviewed at two-year follow-up.

Notes: Inside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. Next to the bars are the differences between treatment and control mean. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

3. Effects on air quality outcomes

The improvement of air quality due to availability of central heating might have an impact on the learning environment. For example, using wood-burning stoves exposes students to fine particle pollution, carbon monoxide, and hazardous air pollutants that can trigger health problems such as asthma and allergies among students and teachers, which might lead to absenteeism and affect academic performance. Indeed, recent studies have established a link between exposure to particulate matter pollution and lower academic achievement (for example, Gilraine and Zheng 2022; Palacios et al. 2022). As part of the building survey, enumerators collected measurements of small particulate matter ($PM_{2.5}$) and coarse particulate matter (PM_{10}) in up to six classrooms per school in February during the two-year follow-up site visits. Using these records, we analyzed air quality for both the "median classroom" in a school (which represents what a visitor should expect if they entered a school building and picked a classroom at random) and for the "max classroom" in each school (the classroom with the worst air pollution levels). As shown in Table III.5, in some schools, the max classroom is an outlier where air quality is far worse than other classrooms in the building. Statistically, these outlier classrooms have a large effect on the average air quality observed across the classrooms visited in each school. For this reason, we are presenting data for the median classroom (rather than the mean across classrooms), to show what air quality looks like in a typical classroom in the building.

Rehabilitation produced a dramatic improvement in air quality measured as exposure to PM2.5 and PM_{10}. Rehabilitation dramatically improved $PM_{2.5}$ and PM_{10} levels in rehabilitated schools (with reductions of -63ppm for $PM_{2.5}$ and -82ppm for PM_{10}, compared to the control group). In rehabilitated schools, the max classroom in each school had average $PM_{2.5}$ and PM_{10} levels of 16 ppm and 35 ppm, respectively. These values were markedly lower than control schools, where $PM_{2.5}$ and PM_{10} levels averaged 79 ppm and 117 ppm, respectively. These findings were not driven by the most extreme classrooms with the worst air quality in each school: we also found similarly large improvements in air quality for the median classroom (Table III.5).

Table III.5. Impact of rehabilitation on air quality outcomes

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Average $PM_{2.5}$ (max classroom)	16.2	79.3	-63.1**	0.000	13.409
Average PM_{10} (max classroom)	35.1	117.0	-81.9**	0.000	16.624
Average $PM_{2.5}$ (median classroom)	10.2	57.5	-47.3**	0.000	11.533
Average PM_{10} (median classroom)	21.0	111.3	-90.3**	0.000	22.854

Source: Building survey administered in 175 schools at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

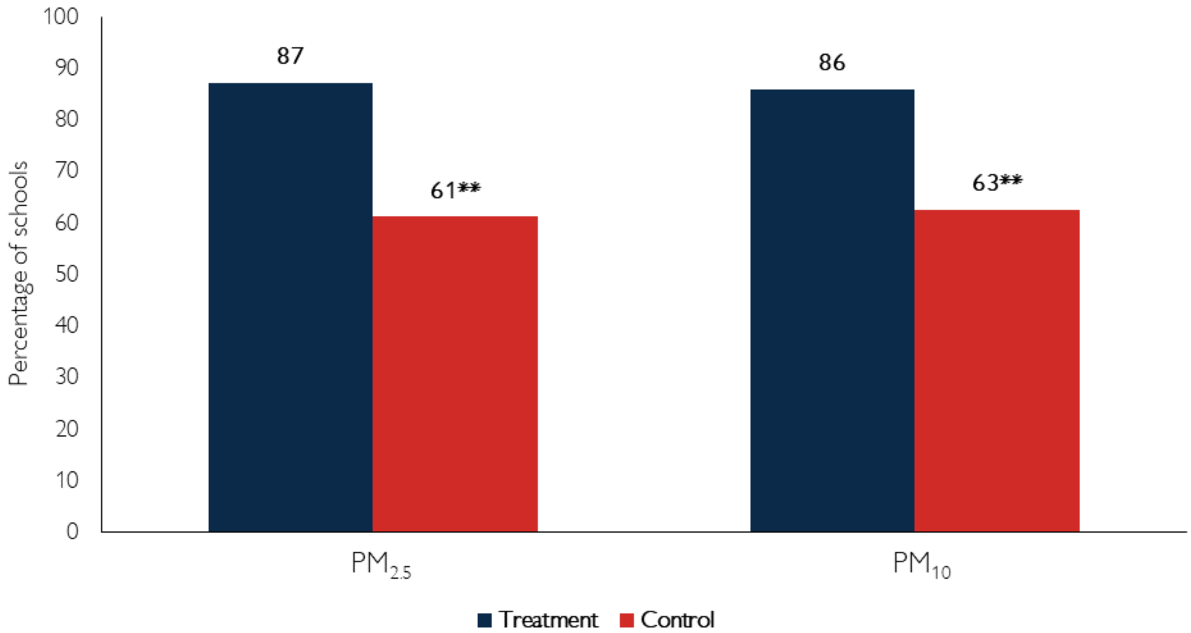
^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Rehabilitation helped schools meet the World Health Organization's interim guidelines for $PM_{2.5}$ and PM_{10} exposure. In 2021, the World Health Organization (WHO) updated its guidelines regarding exposure to $PM_{2.5}$ and PM_{10}. The updated guidelines introduced a long-term goal of reducing $PM_{2.5}$ levels below 5 ppm and reducing PM_{10} levels below 15 ppm. The organization also established nearer-term targets that serve as marker of progress: reducing exposure below 35 ppm for $PM_{2.5}$ and 70 ppm for PM_{10}. The large decline on $PM_{2.5}$ and PM_{10} levels in rehabilitated school resulted in about nine in 10 rehabilitated schools meeting the WHO's interim targets, compared to only six in 10 control schools (Figure III.3). That said, almost none of the rehabilitated schools (or control schools) met the WHO's more stringent, longer-term goal for $PM_{2.5}$ exposure. However, rehabilitation did produce an improvement of 12 percentage points in the percentage of schools meeting the WHO long-term exposure goal for PM_{10}: 20 percent of treatment schools and 8 percent of control schools met the long-term target for PM_{10} exposure.

"This firewood was sometimes wet, sometimes it was windy, the stove pipes didn't fit properly in the windows. The wind was blowing inside, sometimes we were in smoke, we were smoked…, it was horrible, it was hell. Now we now have 21st century heating."

– Director rehabilitated school in Samtskhe-Javakheti

Figure III.3. Percentage of schools meeting World Health Organization interim air quality targets ($PM_{2.5}$ and PM_{10})



Source: Building survey administered in 175 schools at two-year follow-up.

Notes: Inside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. Next to the bars are the differences between treatment and control mean. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

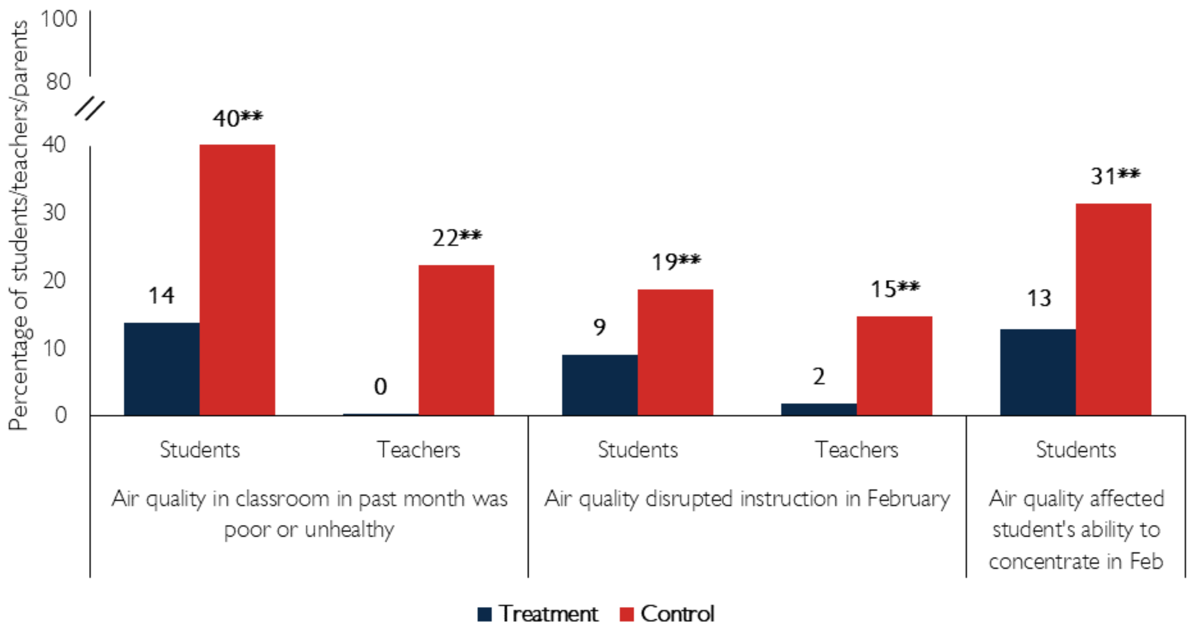
PM = particulate matter.

Students and teachers believed that air quality improvements in rehabilitated schools enhanced the learning environment. Rehabilitation substantially reduced the percentage of students who reported that air quality in the classroom in the past month was poor or unhealthy (a statistically significant impact of about 26 percentage points, Figure III.4). Similarly, the program nearly eliminated teachers' concerns about air quality, and it reduced the percentage of students (-10 percentage points) and teachers (-13 percentage points) who reported that air quality affected instruction during cold weather. Importantly, rehabilitation also produced an improvement of 19 percentage points in the percentage of students who felt that air quality affected their ability to concentrate on schoolwork during the winter (Figure III.4).

> "[Before] when we were using the wood stove, on windy days the smoke was coming in and we were in tears all the time and we were walking in the corridors looking for a place where we could hold lessons. It has changed, and now we do not spend any time on ventilation and stove fuel collection and so on. I used to have to go outside to breathe air before, it does not happen anymore. The lesson process is more organized."
>
> *– STEM Teacher, rehabilitated school in Guria*◢

Qualitative interviews revealed that air quality improvements helped to improve time on task in several ways. In control schools, students and teachers reported that excessive smoke can make it difficult to read or concentrate on written materials, and lessons would sometimes have to be interrupted to search for a different room (or hallway) to avoid the smoke coming from wood stoves. For example, one teacher said they used to only be able to teach for about 30 minutes out of each 45-minute lesson period during bad weather, because of the time spent managing wood stoves (including efforts to avoid smoke, search for alternative teaching locations, or manage issues with the wood supply).

Figure III.4. Impact of rehabilitation on perceived air quality in schools in February



Source: Teacher and student surveys completed by 1,385 teachers and 7,863 students, interviewed at two-year follow-up.

Notes: Inside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. Next to the bars are the differences between treatment and control mean. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

4. Effects on lighting and electrical systems

Rehabilitation greatly improved classroom lighting conditions and addressed lighting-related impediments to reading and concentration. Nearly all the classrooms in rehabilitated schools had functional electric lighting, whereas about half of control schools had at least one classroom without lighting. As shown in Table III.6, the difference of -47 percentage points was statistically significant. This produced a marked improvement in reading conditions; rehabilitation produced large, statistically significant reductions in the percentage of students who reported that lighting issues make it more difficult to read in the classroom (-31 percentage points), read the blackboard (-29 percentage points), or concentrate on classroom activities (-23 percentage points). Similarly, fewer teachers in rehabilitated schools reported that lighting was inadequate for students (6 percent versus 35 percent of teachers in control schools). Qualitative interviews with teachers in control schools revealed that lighting problems were particularly acute in winter months (when there is less natural daylight and the beginning or end of the school day is conducted in darkness).

In addition to the direct benefits from lighting upgrades, qualitative interviews revealed that teachers and students also benefited from improved electrical wiring and access to outlets. Respondents noted that prior to rehabilitation, many schools only had partial access to electrical wiring, meaning teachers were forced to manage with very long extension cords and multiplug adapters that would regularly malfunction or break. For lessons that require access to electric equipment (such as a projector), this led to interruptions when equipment malfunctioned, such as having to pause lessons to find replacement extension cords or relocating the class to hallways or alternative classrooms. The rehabilitation program fixed these problems by installing functioning electrical outlets (in addition to light switches) in every classroom.

> "Previously we did not have electricity at all. We only had electricity in rooms which were used for administrative purposes, because you need internet, you need a computer, and it was transmitted through wires and we could not have electricity in classrooms. And the children could not enjoy it in any way. The internet has improved now because we have Wi- fi all over the school."
>
> *– Director of rehabilitated school in Shida Qartli*◢

Table III.6. Impact of rehabilitation on lighting and its effect on the learning environment

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools					
At least one classroom without working lighting in school	3.4	50.6	-47.2**	0.000	0.058
Students					
Ever have difficulty reading because of lighting	5.3	36.2	-30.9**	0.000	0.027
Ever have difficulty reading blackboard because of lighting	34.9	63.7	-28.8**	0.000	0.023
Feels lighting negatively affected ability to concentrate on schoolwork in February	5.4	27.9	-22.5**	0.000	0.023
Teachers					
Feels lighting is insufficient for students	6.0	35.3	-29.3**	0.000	0.044

Source: Teacher and student surveys completed by 1,385 teachers and 7,863 students, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Rehabilitation addressed problems with classroom equipment. Rehabilitation substantially reduced the percentage of schools where at least one teacher reported a problem with desks, chairs, blackboard, or instructional materials: the impact estimate (39 percentage points) was statistically significant. Similarly, teachers in rehabilitated schools were far less likely than control teachers to report problems with desks

(10 percent compared to 32 percent), chairs (8 percent compared to 30 percent), blackboard (0 percent compared to 25 percent), or instructional materials (0 percent compared to 33 percent). Each of these impacts is statistically significant.

Table III.7. Impact of rehabilitation on classroom equipment

	Treatment (A)	Control (B)	Difference (A-B)	*p*-value	Standard error
Schools					
Percentage of schools with at least one teacher reporting problems with desks, chairs, blackboard, or instructional material in their classroom	45.3	84.7	-39.4**	0.000	0.072
Teachers					
Percentage of teachers who reported having problems in the classroom with the following equipment					
Desks	10.0	32.1	-22.1**	0.000	0.043
Chairs	7.5	30.1	-22.6**	0.000	0.039
Blackboard/whiteboard	0.0	24.7	-24.7**	0.000	0.031
Instructional materials	0.3	33.4	-33.0**	0.000	0.035

Source: Teacher completed by 1,385 teachers from 175 schools, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

5. Effects on sanitation outcomes

The program also delivered significant improvements to sanitary facilities and cleanliness of toilet facilities in rehabilitated schools. Rehabilitation reduced the percentage of schools without a functional toilet by 16 percentage points, a statistically significant impact. About 18 percent of rehabilitated schools did not have a flush toilet (meaning they were using pit latrines), compared to 34 percent of control schools. Moreover, fewer treatment schools (3 percent) had flush toilets that were not functional, compared to control schools (17 percent, results not shown). The program also improved sanitary conditions and cleanliness of toilet facilities in treatment schools. Treatment schools were far less likely than control schools to lack running water for hand washing (22 compared to 42 percent), lack

"[Before the rehabilitation] we had one toilet outside, a wooden toilet. Even if I had brought chlorine and some disinfectants, it would just be wasted because of the poor conditions. In addition, the whole school was relying on that one toilet, so class breaks did not give enough time for everyone to use it. Students used to go out to the toilet during lessons; this was hindering the learning process a lot. Now we have six toilets indoors, everything is well furnished, and the children think their environment is fabulous."

– Director of rehabilitated school in Samegrelo-Zemo Svaneti◢

soap (6 compared to 26 percent), or have an odor in restroom facilities (22 compared to 70 percent). Each of these impacts is statistically significant.

Rehabilitation substantially improved student comfort using sanitary facilities. Rehabilitation produced an improvement of 33 percentage points in the percentage of students who reported “always” feeling comfortable using sanitary facilities. This change was driven by a dramatic improvement in the percentage of students reporting that they “rarely” or “never” felt comfortable using sanitary facilities: for example, about 38 percent of students from control schools reported never feeling comfortable, compared with 12 percent of students in rehabilitated schools[8]. Student focus groups revealed that the changes that were most appreciated by students included ensuring that all toilet facilities were functional, separated by gender, and located inside the school (some control schools use outhouses outside of the main school building). Teachers also pointed out that students in rehabilitated schools require less time to reach toilet facilities, which helps lessons start on time and with fewer interruptions.

Table III.8. Impact of rehabilitation on sanitary facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools without:					
Flushing toilet	17.8	33.7	-15.9*	0.022	0.069
Running water for hand washing	6.4	25.7	-19.3**	0.000	0.055
Soap near toilets or latrines	21.8	42.3	-20.5**	0.003	0.068
Schools with an odor in restroom facilities	22.4	70.4	-48.1**	0.000	0.060

Source: Building survey administered in 175 schools at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

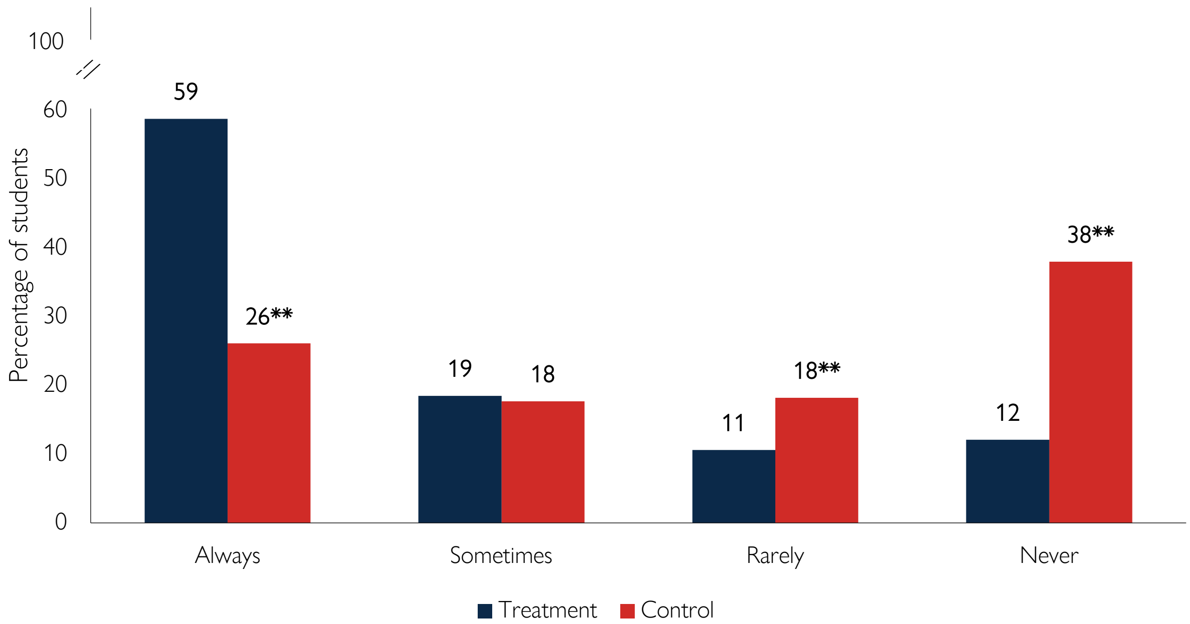
^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Improvements in comfort using sanitary facilities were similar among female and male students. We conducted a subgroup analysis to assess whether the pattern of impacts was different between male and female students, and we found no statistically significant differences. Results for female students were very positive (an improvement of 34 percentage points in the percentage who “always” felt comfortable) and similar to the positive results observed among male students (an impact of 31 percentage points). While the differences between treatment and control students were statistically significant for both male and female students, results for both genders are statistically indistinguishable

[8] Similar differences remain after combining students who reported feeling “always” or “sometimes” comfortable into one group, and combining students who “rarely” or “never” feel comfortable into a second group. Rehabilitation produced an improvement of 34 percentage points in the percentage of students who reported “always or sometimes” feeling comfortable using sanitary facilities. Similarly, there was a substantial improvement in the percentage of students reporting that they “rarely or never” felt comfortable using sanitary facilities: about 56 percent of students from control schools reported “never or rarely” feeling comfortable, compared with 23 percent of students in rehabilitated schools.

from one another (Table III.9).[9] The evaluation's (mixed-gender) student focus groups also did not reveal any notable differences in the ways improved sanitary facilities affected male versus female students.

Figure III.5. Impact of rehabilitation on student comfort using sanitary facilities



Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: Inside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. Next to the bars are the differences between treatment and control mean. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

In focus groups, students in rehabilitated schools asked for further improvements in sanitary-facility maintenance and access to drinking water. Despite overall sanitation infrastructure improvements, in focus groups, the students in rehabilitated schools reported that there are ongoing challenges with keeping upgraded sanitary facilities in good working order (occasional clogged pipes, broken sink faucets, stall doors that fail to latch, and empty soap and paper towel dispensers were each mentioned multiple times). In focus groups, many students, as well as interviewed teachers and school directors, also noted that there is a lack of ready access to drinking water for students. In several schools, there is no drinking water available outside of the faucets in bathrooms, and some students do not feel comfortable with the cleanliness of drinking water from those sources. In other schools, a drinking fountain is available near the school's outdoor sports field, but this location is far from classrooms, leading to water breaks taking time away from instruction.

[9] The subgroup results are very similar if we collapse the "always" and "sometimes" categories into one group and the "rarely" and "never" categories into a second group. For female students there was an improvement of 35 percentage points in the percentage who "always or sometimes" felt comfortable using sanitary facilities. These results were statistically indistinguishable from the positive results observed among male students (an impact of 32 percentage points).

Table III.9. Impact of rehabilitation on student comfort using sanitary facilities, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Students who reported they were:					
Always comfortable using the sanitary facilities	34.3**	0.000	31.1**	0.000	0.264
Sometimes comfortable using the sanitary facilities	1.0	0.579	0.6	0.670	0.867
Rarely comfortable using the sanitary facilities	-9.0**	0.000	-6.2**	0.000	0.103
Never comfortable using the sanitary facilities	-26.3**	0.000	-25.5**	0.000	0.789

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

C. Effects on instructional time, facility use, and school safety

1. Instructional time

The theory of change for the school rehabilitation program assumes that improvements in infrastructure will lead to an increase in “time on task” during the school day. There are multiple ways to interpret and measure this outcome. One possibility is that school rehabilitation could improve student attendance rates (increasing the number of days students are present for learning activities during the school year). Another possibility is that infrastructure improvements (such as lighting, heating, or air quality) could make it possible to increase the amount of uninterrupted time spent on classroom instruction within a school day. This evaluation did not include a systematic classroom observation effort to directly measure minutes of instruction in classrooms. Instead, the study triangulated survey responses from teachers (regarding the number of minutes they recollect spending on instruction each day, and the presence or lack of infrastructure-related interruptions during classroom time) with qualitative interviews exploring how school infrastructure affects teachers’ ability to deliver focused instruction time. The program logic assumes that increased time on task will lead to improved student learning outcomes (and ultimately increases in educational attainment, employment, and earnings).

Rehabilitation did not affect absenteeism rates reported by teachers. By the two-year follow-up survey, we found no differences between treatment and control schools in the absenteeism patterns reported by teachers. In both groups, about 58 percent of teachers reported that between one and five students were absent on an average day in the previous month. Similarly, approximately one-third of teachers in both groups reported that five or more students were absent on an average day in the previous month. We also estimated the percentage of enrolled students absent on an average day and found no differences between treatment and control groups.[10]

However, the relationship between the school infrastructure and absenteeism may have been disrupted by the COVID-19 pandemic. In qualitative interviews conducted in 2022, teachers pointed out that the effects of the COVID-19 pandemic may have eliminated gains in attendance rates that would have otherwise occurred. In multiple schools, teachers (and school directors) said that they believed student attendance rates did improve initially after rehabilitation was completed, but absenteeism has increased markedly during the years of the pandemic, when students were more likely to be held out of school due to the risk of illness. These teachers believed that building improvements did initially produce feelings of safety and well-being among students and families that in turn improved student attendance rates, but this effect on attendance rates may have been overwhelmed by the perceived and actual risks of attending school in person during the pandemic.

Table III.10. Impact of rehabilitation on reported student absences

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers reported students' absences on an average day in the previous month:					
Perfect attendance	12.6	10.2	2.5	0.368	0.028
Between one and four students absent	57.7	58.1	-0.4	0.928	0.043
Five or more students absent	29.6	31.7	-2.1	0.632	0.044
Percentage of enrolled students absent on average day	16.6	17.2	-0.5	0.646	0.012

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

a. Teacher-reported instructional time per day

Rehabilitation also did not affect the amount of time teachers report spending on classroom instruction. Despite lighting, heating, and air quality improvements in classrooms, teachers reported spending a similar amount of time in total on classroom instruction in treatment and control schools. About 36 percent of treatment and 39 percent of control teachers spent, on average, between one and two hours on instruction per day. Similarly, approximately 42 percent of treatment teachers and 36 percent of control teachers spent, on average, three to four hours on instruction per day.

[10] In addition to teacher perception data, for the interim evaluation report the study's data collection teams also conducted student head counts as part of each data collection visit and compared these head counts to the enrollment totals for the school (as reported by EMIS) to calculate attendance rates on the day of the visit. These directly observed attendance rates broadly mirror the reported attendance rates in teacher surveys (Nichols-Barrer et al. 2019).

> There was a period when we used to heat with firewood and each time it was relit, we spent 15 minutes on it. Oh, how many times I used to have to light it and bring firewood from home. The distribution of time has naturally gone better in this regard. The more time that you do not spend on something else, you get to spend on learning, and naturally the result is better
>
> *- STEM teacher, rehabilitated school in Samegrelo-Zemo Svaneti*◢

However, these measures of the total amount of time spent on instruction do not account for teachers' ability to use classroom time effectively. This study's survey data about time-use consists of teachers reporting the number of minutes that they allocated to classroom instruction each day, but it does not account for the extent to which they could use classroom time to meaningfully engage students in learning without distraction or discomfort. To more fully understand time-use, we conducted in-depth interviews with teachers and focus groups with students to understand how rehabilitation changed the way classroom time is used. As discussed above, teachers and students in rehabilitated schools noted that the change from wood-fueled stoves to central heating has helped increase the quality of the time they can spend on instruction (due to more comfortable temperatures and improved air quality). Respondents also consistently noted that upgraded electrical systems and lighting directly improved teachers' ability to use classroom time effectively as well.

Table III.11. Impact of rehabilitation on class time spent on instruction per day

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Class time spent on instruction per day in the month before the two-year follow-up surveys:					
Less than an hour	4.9	3.3	1.5	0.205	0.012
One to two hours	36.1	38.9	-2.8	0.532	0.044
Three to four hours	41.8	35.7	6.1	0.130	0.040
Five or more hours	17.3	22.1	-4.8	0.154	0.034

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

2. Use of science labs

Rehabilitation investments delivered new science laboratories and equipment. Rehabilitation increased the percentage of schools with a science laboratory by 58 percentage points, a statistically significant impact. After rehabilitation, nearly all students (97 percent) noted having a science laboratory in their school, compared to 39 percent of students from control schools.[11] Similarly, as shown in Table

[11] The study's infrastructure assessment teams confirmed that 95 percent of rehabilitated schools had access to a science laboratory at the time of the school visit, which is broadly consistent with the student survey results. While all rehabilitated schools did receive a science lab, a few of these schools fully closed off all access to their science

III.11, rehabilitation increased the availability of science equipment such as microscopes (21 percentage points), heating devices (30 percentage points), chemicals (30 percentage points), lab coats (37 percentage points), protective eyewear (40 percentage points), and beakers (14 percentage points). We also found differences in the availability of projection screens (9 percentage points) and internet access (18 percentage points) between treatment and control schools (likely a result of upgraded electrical systems provided through the program).

Rehabilitation also produced significant improvements in students' exposure to science laboratories, including receiving more science demonstrations and participating in experiments. Rehabilitation dramatically increased the percentage of science teachers using demonstrations or experiments. It was much more common for science teachers in rehabilitated schools to provide science demonstrations (80 compared to 20 percent of science teachers in control schools) and conduct experiments during the school year (77 compared to 14 percent), as shown in Table III.12. Similarly, the program increased the percentage of students who reported that their science teacher "always" or "sometimes" shows demonstrations (by 33 percentage points) or provides an opportunity to conduct experiments (by 27 percentage points).

Table III.12. Impact of rehabilitation on students' exposure to science laboratories

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers					
Always/sometimes shown demonstrations (if teaching science)	80.3	19.6	60.7**	0.000	0.041
Always/sometimes conduct own experiments (if teaching science)	77.3	13.8	63.5**	0.000	0.041
Students					
School has a science laboratory	97.0	39.1	57.8**	0.000	0.036
Students who reported teacher always/sometimes shown demonstrations	56.3	23.8	32.5**	0.000	0.031
Students always/sometimes conduct own experiments (if teachers did demos)	53.1	26.3	26.8**	0.000	0.025
Availability of science equipment (if teachers did demos)					
Equipment not available in science class	0.1	8.1	-8.0**	0.000	0.014
Microscopes	76.7	55.8	20.9**	0.000	0.031
Heating devices	48.6	18.5	30.1**	0.000	0.024
Chemicals or other materials for experiments	64.7	34.8	29.9**	0.000	0.034
Lab coats	45.8	8.5	37.3**	0.000	0.027
Protective eyewear	50.7	10.8	39.8**	0.000	0.026
Beakers	64.3	50.5	13.8**	0.000	0.026
Other science equipment	1.3	1.1	0.3	0.424	0.003
Availability of electronic equipment (if teachers did demos)					

labs at the time of the study team's data collection visit due to pandemic-related social distancing rules, as discussed further below.

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Computer	48.2	48.7	-0.5	0.876	0.032
Internet access	40.7	24.6	16.0**	0.000	0.031
Projection screen	37.2	28.5	8.8**	0.002	0.028
Television	7.0	7.3	-0.3	0.901	0.023
Other electronic equipment	1.6	1.2	0.4	0.314	0.004

Source: Teacher surveys completed by 532 science teachers and 5,444 students who received science demonstrations, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

However, in qualitative interviews, science teachers identified several barriers that have made it difficult to access science laboratories as often as they would like. While rehabilitation clearly succeeded in delivering upgraded science labs that are used to provide demonstrations and experiments, qualitative data indicate that there have been important barriers to accessing these facilities. To a certain extent, this is a sign that the program succeeded in providing a set of facilities that are receiving a high level of demand from teachers (indicating also that teachers felt comfortable integrating lab facilities into their curriculum and lesson plans). On the other hand, the fact that teachers reported that they were not able to use the labs as often as hoped indicates that science laboratories have some unfulfilled potential to affect science instruction even further. These barriers to lab access fall into three categories: pandemic-related disruptions, challenges related to laboratory capacity and supplies, and challenges with heating science labs in winter.

While the COVID-19 pandemic could not have been predicted by the program, it had a profound effect on science-lab access during the latter part of this evaluation's data collection period. Social distancing regulations in schools effectively eliminated access to science laboratories for the latter half of the 2019–2020 school year (when schools were closed) and much of the 2020–2021 and 2021–2022 school years (when schools were required to maintain several feet of separation between students—this configuration is not possible in science labs set up for small groups of students to gather around lab benches).

> "I need a laboratory, I need to use the laboratory because I teach physics and math, but unfortunately, there is only one laboratory for all of the science subjects, and it is not enough for all the classes […] I try to overcome this […] But I was told I wasn't allowed to take any equipment out of the laboratory and thus everything is in there gathering dust, untouched. So, I can't take the equipment out to my rooms and I can't use the laboratory either, it would be nice if we had a solution to this problem"
>
> *.- STEM teacher from rehabilitated school in Samtskhe-Javakheti*◢

Apart from the pandemic, teachers also reported capacity issues limiting their ability to use the science labs as intensively as they had hoped. In multiple rehabilitated schools, respondents reported that there is insufficient lab space: science teachers report that they only have one lab room that must be shared across three science subjects (biology, chemistry, and physics), creating logistical difficulties due to the need to set up the lab for different subjects between class periods. In some of these schools, teachers also reported that the lab is too small to fit all the students in their classes, although these responses may have been affected by the temporary social-distancing requirements of the pandemic. An additional set of concerns relates to keeping an adequate stock of lab supplies; in multiple schools, chemistry teachers noted that the labs did not have an adequate stock of reagents for the more-complex experiments they would like to perform with upper-grade students, and students noted that they did not have enough of various types of equipment for every student to practice using them.

A final set of concerns about science-lab access related to heating these rooms adequately during winter months. Because rehabilitated schools use central heating systems, heating rooms adequately required unobstructed access to heating vents or radiators located along classroom walls. Respondents in multiple rehabilitated schools reported that the need to install large cabinets in the science labs ended up blocking heating vents in the room and preventing them from working properly. The result was that science labs in these schools have been too cold to use comfortably during the coldest months of the year; addressing this issue would involve relocating or removing cabinets and equipment (for example, storing some items in a different part of the school), so that the heating system can fully function.[12]

3. Use of recreational facilities

Rehabilitation in treatment schools increased usage of indoor recreational facilities. Rehabilitation increased the percentage of students reporting that they used an indoor gym at least once in an average week by 6 percentage points, a statistically significant impact (Table III.13). That said, most schools had at least some type of indoor gym prior to rehabilitation; the program modestly increased the availability of indoor gyms (by 7 percentage points), and the impact was only statistically significant at the 10 percent level. We also found no differences in the percentage of students reporting that they used outdoor recreational spaces. Student focus groups also revealed that rehabilitation provided particularly important improvements to indoor-gym safety, especially with respect to replaced flooring that fixed dangerous uneven or missing wooden floorboards that students could trip over while running. Respondents reported that they feel much safer using the gym after rehabilitation, and some students in treatment schools reported that the improved gym infrastructure has increased their interest in sports.

Table III.13. Impact of rehabilitation on use of recreational school facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Availability of indoor gym	86.3	79.3	7.0	0.080	0.040
Student reported using at least once in an average week					
Indoor gym (if available)	92.5	86.5	6.0*	0.012	0.024
Outdoor recreation area (if available)	68.8	65.9	2.9	0.498	0.043

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

[12] The study's quantitative infrastructure assessments did not assess the adequacy of central heating in science labs on a separate basis from other classrooms, so we cannot directly observe how often this issue occurred across all rehabilitated schools.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

4. School safety

Students, parents, teachers, and school directors felt that rehabilitation improved safety in schools. Rehabilitation substantially improved the percentage of respondents who felt safe in the school, with statistically significant impacts of 43 percentage points for students, 35 percentage points for parents, 35 percentage points for teachers, and 41 percentage points for school directors. As shown in Figure III.6, more students, parents, teachers, and school directors from treatment than control schools reported that school facilities were safe. About 90 percent of students and 99 percent of teachers in the treatment group reported feeling very safe in the classroom, compared to 51 percent of students and 66 percent of teachers in the control group. Similarly, most students (87 percent) and all teachers from treatment schools also reported feeling very safe using the stairwells, whereas in control schools, only 43 percent of students and 64 percent of teachers reported feeling very safe.

In qualitative interviews, students and teachers cited many aspects of rehabilitation that played a large role in helping to make the school feel safer. Respondents reported that rehabilitation largely removed the risk of ceiling material falling or leaking, gaps in flooring and stairways, or broken windows. Two teachers mentioned that moving toilets inside the school building reduced feelings of discomfort among students, especially younger students, who had difficulty using outdoor facilities. Improved heating systems also played a large role in making the school feel safer. All types of respondents mentioned that before the school was rehabilitated, the low temperature in classrooms could cause students and teachers to fall sick and that smoke from wood stoves was a threat to respiratory health. A few teachers also reported that wood stoves had the potential to burn students; students in one focus group also discussed how they feel protected from fire now that the school has an emergency door and fire extinguisher, and they are not lighting wood stoves throughout the day.

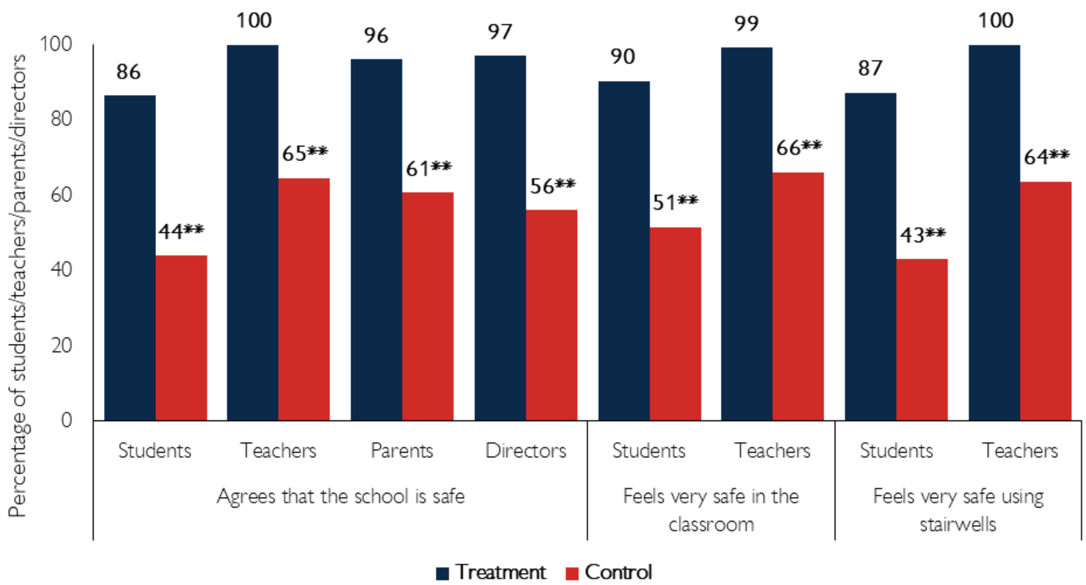
"Damaged ceilings hinder the learning process, especially in rainy weather, and it is not safe. [In one classroom] things were falling down from the ceiling so much that the school management decided to stop using the room. The flooring is completely failing in some classes: if a student is walking by and forgets about that danger, they may fall in a hole and break their leg."

– Language arts teacher, comparison school in Imereti

"I am in a safe environment, and we are not so tense that the child might be in danger. This is the first time when children are safer and when I'm in peace as a teacher. I contribute more to the lessons."

–STEM teacher, rehabilitated school in Imereti◢

Figure III.6. Impact of rehabilitation on perceived safety



Notes: Outside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

D. Effects on learning outcomes

To measure the effects of school rehabilitation on student achievement, we examined test scores from three subject tests (Georgian, math, and science) designed and administered by Georgia's NAEC for students enrolled in grades 9, 10, 11, and 12 during follow-up visits to study schools.[13] To make the results comparable across grade levels, we converted scaled test scores to *z*-scores within the sample for each subject and grade (separately for grades 10, 11, and 12). A *z*-score value indicates how far a student scored from the average, using standard deviation units to define the scale. For example, a *z*-score of zero means a student scored at the 50th percentile, whereas a *z*-score of 1.0 would represent a score at the 84th percentile.

Student test scores in language, math, and science were very similar in treatment and control schools. For language and math, we observed *z*-scores that were 0.02 standard deviations higher in treatment schools compared to control schools, equivalent to moving a student's score from the 52nd to the 53rd percentile. In science, on the other hand, we found a difference of 0.07 standard deviations in favor of the control group, equivalent to decreasing a student's rank from the 56th to the 54th percentile in the study sample. However, none of these differences were statistically significant. In other words, for the overall sample, the evaluation did not find strong evidence that school rehabilitation increased test scores in upper-secondary grades two years after rehabilitation was completed. (Table III.14).

[13] NAEC staff administered tests in grades 9–11 in one-year follow-up visits and in grades 10–12 in two-year follow-up site visits. The program's impacts on test scores after one year can be found in Appendix B.

Table III.14. Impact of rehabilitation on student test scores across grades 10, 11, and 12

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Language	0.06	0.04	0.02	0.81	0.066
Math	0.05	0.03	0.02	0.799	0.074
Science	0.08	0.16	-0.07	0.295	0.071

Source: Administrative data from the Education Management Information System.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

While all of these effects are close to zero, the more-negative impact of school rehabilitation in science during the study period could be related to restrictions on in-person learning activities during the COVID-19 pandemic. As described above, over two school years in the study's follow-up period (2020-2021 and 2021-2022) social distancing requirements prevented teachers and students from accessing science labs even when the school building was open. As a result, social distancing rules may have been more harmful to science instruction in rehabilitated schools (where teachers and students were looking forward to using improved labs) relative to control schools (many of which lacked access to labs in the first place), and contributed to the Activity's more-negative effect on science learning outcomes during this period. We explore the effects of the pandemic in more detail below.

Broadly speaking, it is likely that the COVID-19 pandemic may have disrupted the relationship between rehabilitation and student learning outcomes in important ways. The pandemic produced the following large-scale changes in the Georgian education system in the 2019-2020, 2020-2021, and 2021-2022 school years:

- Schools in Georgia were entirely closed for most of March 2020, and they began offering remote (online) learning options from April 2020 until the end of the 2019-2020 school year.
- At the beginning of the 2020-2021 school year, about a third of school buildings remained closed and a larger proportion of schools enacted intermittent closures during the school year in response to local COVID-19 outbreaks.
- By the end of the spring 2021 semester, nearly all schools in Georgia were offering in-person learning again. However, for all of the 2020-2021 and 2021-2022 school years social distancing restrictions were enforced in schools that were open for in-person learning activities; these rules required students and staff to mask and remain several feet apart throughout the school day.

Because school rehabilitation took place between 2016 and 2019, the evaluation measured test score outcomes prior to the pandemic for some schools and after the pandemic began for other schools. While the study's sample size is not large enough to carry out school-level subgroup analyses precisely, we carried out an exploratory analysis that tested whether the effects of rehabilitation on learning outcomes appear different in the schools where data collection was completed prior to the pandemic. This is important to examine because students assessed before the pandemic experienced two years of uninterrupted exposure to improved school buildings, whereas students assessed after the pandemic experienced multiple extended school closures (with schools closed for the entire spring 2020 semester and then intermittently closed for portions of the 2020–2021 school year).

In language and math, rehabilitation appears to have had a positive effect on learning outcomes before the pandemic, but these effects were absent for schools assessed after the pandemic. While the patterns in Figure III.7 are highly suggestive, the sample sizes in pre-pandemic and post-pandemic groups of schools are not large enough to detect if the impacts in the two periods are statistically significant. In addition, the *difference* between pre- and post-pandemic impacts was only marginally statistically significant (at the 10 percent level) for language test scores and not significant for math test scores. That said, if the pre-pandemic impacts of rehabilitation had held for the entire sample, impacts of that magnitude would have represented a meaningful amount of learning growth. For the language exam, the pre-pandemic impact of 0.13 standard deviations is equivalent to about 6.8 months of learning for upper-secondary students. Similarly, the pre-pandemic impact on math scores (0.10 standard deviations) is equivalent to 5.3 months of learning.[14] In science, the pre-pandemic effect of rehabilitation was close to zero, and the impact on science scores became more-markedly negative after the pandemic. The difference in impacts on science scores across the two periods was not statistically significant. Evidence from qualitative interviews, discussed below, suggests several potential explanations for why the effects of school rehabilitation could have become more negative during the pandemic.

Qualitative data suggest that school closures during the pandemic were especially disappointing to students and teachers in rehabilitated schools. School closures affected the learning process in both rehabilitated schools and comparison schools: regardless of whether the school had been rehabilitated, students, teachers, and directors reported that much of instruction over the past 2.5 school years since the start of the pandemic had been interrupted or curbed by shifts to full or partial remote learning and social distancing restrictions. However, the "opportunity cost" of the shift to remote learning appears to have been especially great for students and teachers in the treatment group, who had been looking forward to accessing newly upgraded school buildings and science labs. Students and teachers both reported that they were greatly disappointed to have lost access to the newly renovated infrastructure during the pandemic: many mentioned how painful it was to learn of the shift to remote learning very shortly after their renovated school building opened. In contrast, the shift to remote learning may not have been as harmful to students in control schools who already had serious concerns about the comfort and adequacy of the learning environment in their schools. In other words, it is plausible that there was a negative interaction effect between (a) receiving an upgraded school building and (b) having access to that building taken away suddenly during the pandemic.

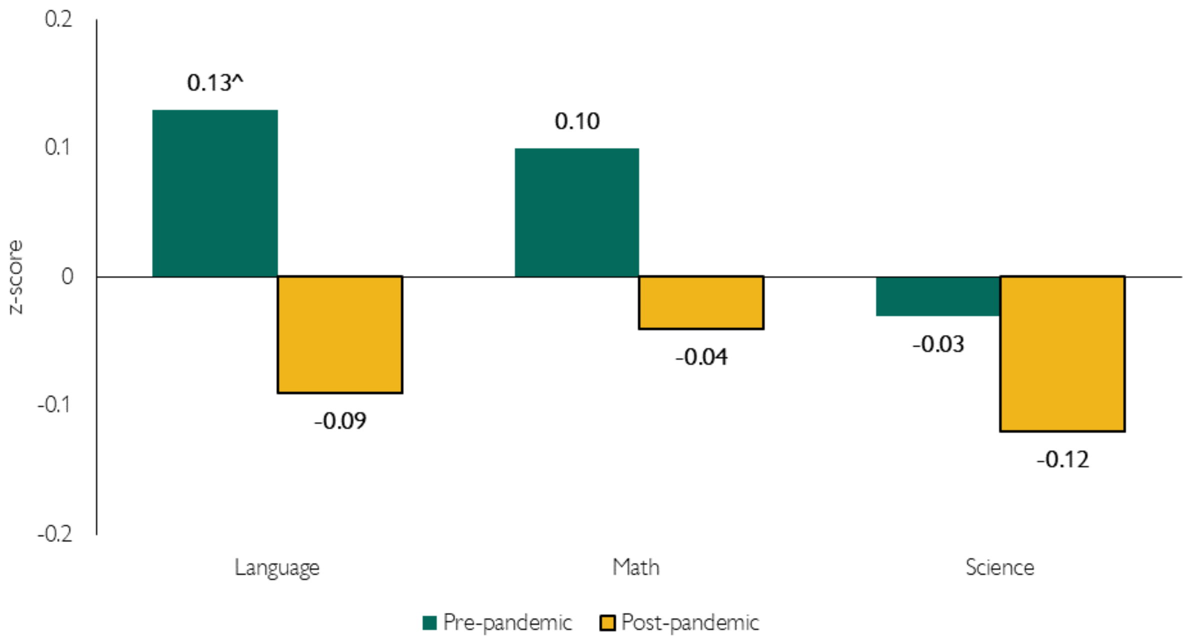
> "I go back to the fact that at first when we came everyone was happy, everyone was interested in even going to the lab, entering computer rooms, working directly with the computer, and then the pandemic prevented us from attending school. Not because it was an uncomfortable environment, it was because of the pandemic."
>
> *– STEM teacher in rehabilitated school in Imereti*◢

Interviews also suggest that difficulties with science lab capacity and access may help to explain why the program did not improve science learning outcomes. As discussed above, teachers, directors and students in rehabilitated schools all reported that they have not been allowed to use science

[14] To calculate months of learning, we used the average annual gain (in effect size units) from a nationally normed test using a vertical scale to compare growth in test scores across grade levels, as reported in Hill et al. (2008).

laboratories for major portions of the pandemic due to the government's social distancing regulations (which prevent small-group work around science-lab tables). Science teachers reported that it was particularly frustrating to have completed training and planning to prepare for using for these labs and then have their lab access taken away. In addition, even before the pandemic, respondents noted access difficulties related to sharing a single lab room across all science subjects and the fact that labs are not large enough to accommodate larger class sizes. In some rehabilitated schools, the science labs were also not adequately heated in winter months (due to equipment and cabinets blocking heating vents or radiators).

Figure III.7. Impact of rehabilitation on student test scores across grades 10, 11, and 12



Source: Administrative data from the Education Management Information System.

Notes: Outside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

E. Longer-term effects of school rehabilitation

In response to the disruptions of the pandemic, the evaluation included a final set of site visits in all study schools in 2022 to assess school infrastructure and learning outcomes at a single point in time. This data collection round served two purposes. First, it provides a snapshot of key outcomes in all rehabilitated school at one point in time, when all schools had experienced the disruptions of the COVID-19 pandemic (that is, the 2022 data collection round included schools rehabilitated in 2016 and 2017, where the pandemic occurred in the third, fourth, and/or fifth follow-up year after the program was completed). Thus, it enables us to look at post-pandemic outcomes in schools that were rehabilitated before the onset of COVID-19 and may have interacted with the pandemic differently than those more recently rehabilitated. Second, this data collection round provides an opportunity to investigate the program's longer-term impacts (among the roughly 60 percent of schools where this visit occurred 3–5 years after rehabilitation) and compare the findings to the (roughly 40 percent of) schools that were only in use for two years after rehabilitation.

"The school infrastructure helps, but the pandemic really hindered us. Two years ago, we were actively using these labs, that were arranged for us by MCA-G. But now we actually do it very rarely, because it is forbidden to work in a classroom without distancing… In general, all these isolations have brought us very bad results …there really was a lot more progress and more motivation before the pandemic than in last few years."

– STEM teacher in rehabilitated school in Shida Qartli◢

In 2022, rehabilitation continued to have positive impacts on school building quality and the learning environment. Using data from the follow-up visits in 2022, we found a very similar pattern of impacts on school infrastructure: rehabilitation continued to produce a dramatic improvement in overall building quality, heating systems and heating-system usage, air quality, and sanitary facilities. We also found that the impact estimates for infrastructure outcomes in schools receiving a longer-term follow-up visit in 2022 (schools that were visited more than two years after rehabilitation was completed) remained very similar to the impact estimates in schools that only received a two-year follow-up visit in 2022. (See Appendix D for detailed results.)

As of 2022, rehabilitation may have produced positive effects on student test scores in language, math, and science, but the changes were not statistically significant. As with the two-year follow-up data, we did not find any statistically significant impacts of the program in language, mathematics, or science in 2022. However, unlike the two-year follow-up, the impact estimates were positive in each of these three subjects. As shown in Table III.15, for the language exam we observed a difference of 0.07 standard deviations in favor of treatment group, but the difference was not statistically significant. Math and science scores also increased by .05 standard deviations, but these effects were also statistically indistinguishable from zero.

We also found suggestive evidence that rehabilitation may produce more positive learning benefits after the second follow-up year. The 2022 data collection provided an opportunity to examine longer-term outcomes in schools that were rehabilitated earliest and compare their result to schools completed more recently. Specifically, we conducted a subgroup analysis that tested whether the effects of the program were different in the schools receiving a two-year follow-up in 2022, compared to schools where

2022 represents a longer-term follow up of 3–5 years after rehabilitation.[15] We found suggestive evidence that the program may produce larger impacts on learning outcomes after the second follow-up year (Figure III.8): across all three subjects, the impact estimates were negative or close to zero in schools that were in their second follow-up year in 2022, but positive (and larger in magnitude) in schools that were in their third, fourth, or fifth follow-up year. Although none of the differences between these subgroups were statistically significant, the sample sizes in the study are only large enough to detect a very large shift in learning outcomes for subgroups of schools. If it turns out that the effects on learning measured during longer-term follow-up visits are more representative of the program's true effects, these effect sizes (0.12 standard deviations in language, 0.11 standard deviations in math, and 0.11 standard deviations in science) would represent an educationally meaningful boost in student achievement. For example, an effect size of 0.10 standard deviations would be roughly equivalent to 5.2 months of learning in language and 7.1 months of learning in math for students in grade 10.

Table III.15. Impact of rehabilitation on student test scores in grades 10 and 12 in 2022

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Language	0.03	-0.04	0.07	0.405	0.079
Math	0.02	-0.03	0.05	0.525	0.082
Science	0.01	-0.04	0.05	0.550	0.081

Source: Administrative data from the Education Management Information System.

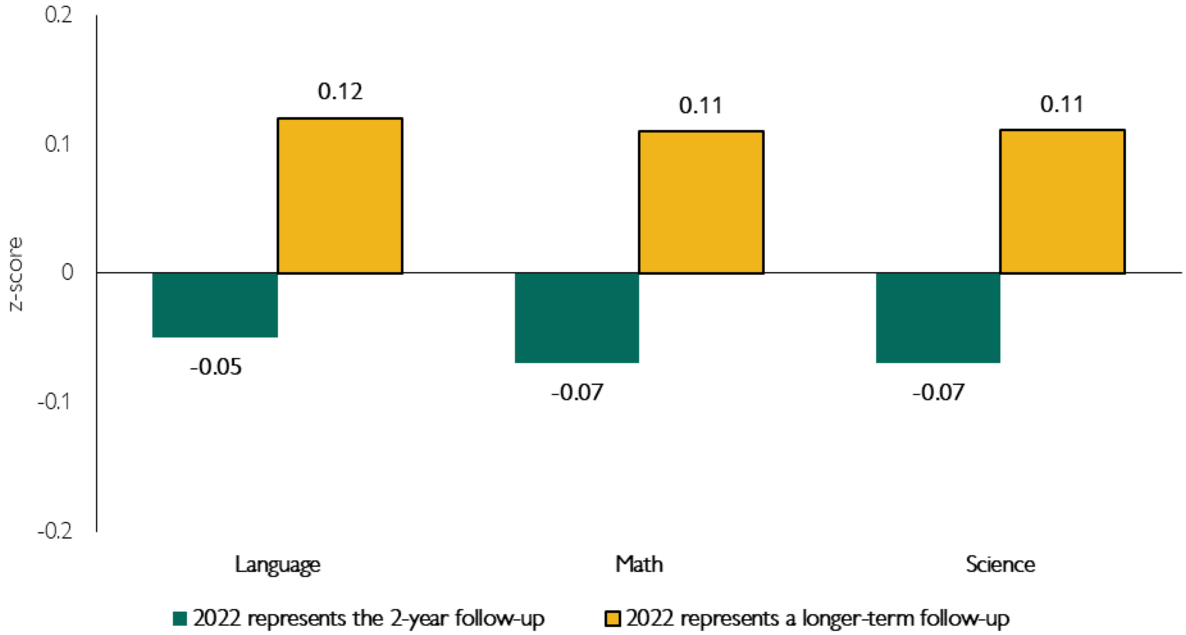
Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

One possible interpretation of this pattern of findings is that changes in the learning environment might require more time to affect learning outcomes, especially considering the pandemic and its disruptions to school-building access in the 2019–2020 and 2020–2021 school years. If so, this would mirror the findings from MCC's investments in school construction in Niger, where the IMAGINE program constructed new school buildings in communities that already had an existing school of some kind. The IMAGINE evaluation found few short-term impacts on enrollment or achievement outcomes (Dumitrescu et al. 2011), but substantial and larger impacts seven years after the program was implemented (Bagby et al. 2017).

[15] Importantly, the program was implemented in phases that were grouped by region within Georgia. As a result, the subgroup analysis discussed here cannot rule out the possibility that the first schools to be rehabilitated may have had a different pattern of impacts than later schools due to factors that are specific to the "Phase I regions" where rehabilitation occurred first (Mtskheta-Mtianeti, Racha-Lechkhumi and Kvemo Svaneti, Samtskhe-Javakheti, and Shida Kartli).

Figure III.8. Impact of rehabilitation on student test scores in 2022



Source: Administrative data from the Education Management Information System.

Notes: Outside the bars are the two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

F. Impacts on enrollment and school administration

1. Impact estimates on enrollment and graduation rates

One possible effect of school rehabilitation could be changes in enrollment levels. Schools with better conditions may attract new students from surrounding schools; the Georgian education system allows families to select a school of their choosing, which could lead to transfers into newer school buildings over time. In addition, rehabilitation investments could help to discourage students from dropping out of school, maintaining enrollment rates at higher levels over time. Under Georgia's per-pupil funding system, increasing enrollment would translate directly to additional operational resources for school directors; on the other hand, accommodating more students could also put pressure on class sizes, instructional resources, or (in extreme cases) require a school to move to a double-shifted schedule where some students attend for the first part of the day and others attend for the second part of the day.

In addition to enrollment, improvements in educational attainment are also directly relevant to the program's theory of change (because they may ultimately be related to increased lifetime earnings). We analyzed student enrollment, dropout, and graduation rates using administrative data provided by EMIS.

Rehabilitation appears to have increased school enrollment by about 10 percent. By the two-year follow-up, treatment schools had an average of 50 more students compared to control schools, equivalent to an increase of 12 percent (Table III.16). Due to the study's random assignment design, treatment and control schools had very similar enrollment levels prior to rehabilitation; as a result, these enrollment differences at the two-year follow-up represent the direct impacts of rehabilitation on total enrollment. While the difference is not statistically significant in our sample, there is some suggestive evidence that enrollment is increasing especially rapidly in early grades. Specifically, we found that rehabilitation produced a 26 percent increase in enrollment in grade 1 (marginally significant at the 10 percent level). The number of students enrolled in grades 2 to 12 is also higher in treatment compared to control schools, but the differences are smaller and not statistically significant.

If the substantial increase in grade 1 enrollment continues with each new cohort entering in subsequent years, it is possible that overall enrollment at these schools could grow larger over time. Indeed, in the subset of schools where it is possible to observe the impacts of school rehabilitation after three years or four years, there is evidence that enrollment levels have been continuing to increase with each successive cohort entering the school.[16]

Qualitative interviews also suggest that these enrollment increases in rehabilitated schools may continue to grow over extended periods. In 2022, the evaluation conducted qualitative interviews with school directors in a subset of schools that received rehabilitation five years earlier. At these schools, directors consistently noted that enrollment had increased noticeably and caused some challenges related to classroom space: one school had to introduce a double-shift schedule to accommodate the influx of students. In another school, the director noted that the number of classrooms became insufficient, so they repurposed assembly halls, gyms, and teacher offices to support instruction in a larger number of classrooms.

Rehabilitation did not affect dropout rates or graduation rates. As shown in Table III.14, we did not find differences between treatment and control schools on dropout rates in grades 8 to 12. We also did not find differences with respect to grade-progression rates (not shown). Likewise, grade 12 graduation rates were also very similar in rehabilitated schools and control schools. Graduation rates for both groups were quite high (approximately 96 percent), suggesting there may not have been a meaningful opportunity to improve educational attainment further in these schools during the study period.

[16] We estimated impacts on school enrollment three and four years after school rehabilitation for a subsample of 126 schools where longer-term follow-up data is available. By the three-year follow-up, rehabilitation produced a 29 percent increase in enrollment in grade 1 (marginally significant at the 10 percent level), as well as a statistically significant 36 percent increase in enrollment in grade 2 (p-value of 0.03). By the four-year follow up, rehabilitation produced similar increases in enrollment in grades 1 and 2 (marginally significant), as well as a statistically significant 36 percent increase in enrollment in grade 3 (p-value of 0.03). After four years, rehabilitation also produced a small but statistically significant decline in grade 9 dropout rates (an impact of -2 percentage points, with a p-value of 0.04).

Table III.16. Impact of rehabilitation on enrollment by two-year follow-up

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Average number of students enrolled per school (2-year follow-up)					
Grade 1	48	38	10^	0.055	5.180
Grade 2	46	40	6	0.239	5.419
Grade 3	45	39	6	0.324	5.651
Grade 4	43	40	3	0.638	5.484
Grade 5	39	35	5	0.336	4.931
Grade 6	38	36	2	0.693	4.553
Grade 7	35	32	4	0.412	4.366
Grade 8	36	34	2	0.543	4.088
Grade 9	37	33	4	0.327	4.069
Grade 10	35	33	8	0.496	3.513
Grade 11	33	31	3	0.376	3.274
Grade 12	32	28	3	0.282	3.133
All grades (1 through 12)	**468**	**418**	**50**	**0.326**	**50.58**
Dropout rate per school					
Grade 8	0.8	1.2	-0.4	0.444	0.005
Grade 9	1.6	3.1	-1.5	0.157	0.011
Grade 10	5.7	5.3	0.4	0.697	0.011
Grade 11	2.3	2.9	-0.6	0.428	0.007
Grade 12	1.6	1.3	0.2	0.683	0.006
Graduation rate (grade 12)[a]	**96.1**	**95.8**	**0.4**	**0.726**	**0.010**

Source: Administrative data from the Education Management Information System.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

[a] Graduation rates were estimated using administrative data from one year after school rehabilitation because data from the two-year follow-up was not available when we made the request.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

2. Impact estimates on school administration outcomes

Rehabilitation nearly doubled the cost of heating the school building in winter, because the costs of central heating are higher than operating wood stoves. Principals from renovated buildings reported higher costs for operating and maintaining school infrastructure, in comparison with directors of control school. Reported heating costs in February in treatment schools were 1,728 GEL (equivalent to 652 USD) higher than the costs reported by school directors in the control group. Directors from rehabilitated schools also had higher electricity costs in February compared to directors in control schools. However, principals from both groups reported similar water costs. We also did not find an impact on the percentage of directors who reported being able to pay for school utilities: there was no effect on the percentage of directors who reported that the school budget was sufficient to pay for maintenance and education activities, and a clear majority of directors (over 70 percent, in both treatment and control schools) reported that they were only "rarely" or "never" unable to pay for school utility expenses (Table III.17). This indicates that the overall utility cost increases following rehabilitation did not overwhelm the school's capacity to pay for utilities on a consistent basis.

Table III.17. Costs incurred between baseline treatment and control schools

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Heating costs for the month of February (in Georgian lari)	3,654	1,926	1,728**	0.000	292.2
Electricity costs for the month of February (in Georgian lari)	579	337	242**	0.000	50.5
Water costs for the month of February (in Georgian lari)	163	140	23	0.532	37.4
Directors who reported school budget was sufficient to pay maintenance costs and education activities	62.8	66.9	-4.1	0.599	0.08
Directors who were unable to fully pay for school utilities (water, electricity, heat, etc.)					
Always	12.3	19.0	-6.7	0.257	0.059
Sometimes	16.1	5.6	10.4*	0.034	0.049
Rarely	10.5	9.3	1.2	0.809	0.050
Never	61.2	66.2	-5.0	0.527	0.079

Source: School director surveys completed by 163 directors, interviewed at two-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

88 out of 95 treatment schools were rehabilitated, and intention-to-treat impact estimates for all treatment schools can be found in Appendix A.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Qualitative data suggest that directors have been able to manage routine operating costs, but funding major repairs is a serious challenge as the infrastructure ages. In 2022, the evaluation collected interview and focus group data from school directors, teachers, and students in a subset of the first tranche of schools to be rehabilitated, providing additional insights about how upgraded infrastructure was maintained over five years after rehabilitation was completed. Most directors reported that they have learned to manage the increased utility costs associated with central heating and do not struggle to pay for it (in part due to the increased revenues provided by enrolling more students into the school). Some directors have also negotiated payment plans enabling them to smooth costs over time (for example, pay some of the winter heating expenses in the summer). However, when major repair needs arise, it remains a major challenge to address the issue. For example, in one school the central boiler malfunctioned, and they have not (thus far) been able to afford repairs. In another school, the wooden floor in the gym was damaged by a water leak, and they lack the funds to fix it.

Five years after rehabilitation, respondents also reported new challenges related to keeping heating systems and sanitation facilities in proper working order. In these long-term follow-up visits, teachers and students reported some increasing problems with sanitation facilities (including such issues as broken flush-toilets, doors that do not close or latch, or problems with low water pressure). One school director said that they must repair the toilets approximately once every two months. Respondents also continued to cite a lack of drinking water as an issue for most students: students say that they just don't feel comfortable drinking out of the bathroom faucets, and they wished that access to cleaner, filtered water from drinking fountains had been part of the rehabilitation program.

In several schools, students and teachers (but not school directors) noted that they were increasingly aware of some difficulties related to heating the school evenly. These respondents complained that the placement of radiators or heating vents does not heat the school evenly (while some classes and corridors are warm, others are cold). Typically, teachers reported that these issues arose because an inadequate number of radiators (or an undersized heating supply pipe) were placed in larger rooms and hallways. In multiple schools, respondents also mentioned that heating science labs continued to be a challenge five years after rehabilitation. Because equipment cabinets in the labs were placed in front of the heaters, these rooms can remain very cold in the winter and difficult to use.

IV. Findings for the TEE Activity

In this chapter, we present the evaluation's final set of findings for the IGEQ Project's nationwide TEE activity. Through the TEE activity, which concluded with the end of the Compact in mid-2019, all of Georgia's grade 7–12 teachers received a series of training modules focused on student-centered instruction, and school directors received a set of related training modules related to teacher observation and school management. Among other topics, the training sought to improve teachers' ability to engage in professional development, develop high-quality lesson plans and assessment strategies, and use instructional practices that foster critical thinking and engage students with material targeted to varied ability levels. Because of its ambitious scope, the TEE activity used a phased implementation schedule, rolling out training to multiple cohorts of teachers over three years. The training sequence consisted of multiple modules (five modules for directors and four for teachers), with each module lasting two to five days. For teachers, the training sequence was held over the course of about one year for each cohort. For school directors, the training was delivered over two years.

Mathematica produced an interim evaluation report for the IGEQ Project in October 2019 (Nichols-Barrer et al. 2019) that described initial teacher and school director outcomes following the training sequence, together with an overarching performance evaluation assessing how the program was implemented. Some of the key findings from the interim evaluation were as follows:

- **The TEE activity succeeded in implementing trainings on a nationwide scale.** In total, the Activity succeeded in holding a sufficient number of training events to offer it to Georgia's whole population of school directors (about 2,000) and all of Georgia's upper-grade teachers in the subjects of science, mathematics, English, and geography (about 18,000 teachers in total). Directors received the training sequence over the course of two years (in a single cohort). Attendance rates at the trainings were generally high. Although school directors completed the full training sequence at a higher rate (93 percent) than teachers in the first cohort (82 percent), a large majority of both groups attended at least one training session, and nearly all of the trainees felt positively about the training experience.
- **After training, teachers showed a pattern of improvements in their knowledge of student-centered instruction strategies.** The interim analysis showed that trained teachers became more confident in their ability to teach higher-order thinking skills and promote cooperation through group work. Trained teachers were also more confident in their ability to use lesson plans that enable differentiated instruction for students with different abilities, use formative assessments in the classroom, and create an equitable environment for girls. Each of these findings represents a statistically significant difference between the trained teachers and the matched comparison group of teachers who had not begun the training sequence; the differences represented increases of 6 to 8 percentage points on knowledge and confidence indices collected in teacher surveys.
- **However, immediately after training we did not find consistent evidence of changes in teachers' classroom practices.** This was expected by program implementers, who designed the TEE activity to encourage changes in teaching practices over longer periods of time. The interim analysis suggests that the training did not change the classroom practices used by trained teachers in the initial period after the training sequence was completed. This finding from the study's matched comparison group analysis was also corroborated by results from surveys of trained teachers, classroom observations (with a small sample of trained teachers), and student surveys (with a convenience sample of students attending classes with trained teachers), all of which showed substantial room for improvement in

teachers' use of the practices encouraged in the training sequence. For example, only 6 percent of students reported that they engage in daily collaborative group work, 16 percent of students reported that they consistently receive the kind of short and informal assessments encouraged by the training (formative assessments), and 10 percent of teachers reported using lessons with differentiated instruction daily. On the other hand, other teaching practices were relatively strong: classroom observations, for example, revealed that teachers were effective in keeping students engaged on instructional tasks and that teachers only rarely use passive instruction techniques (such as asking students to copy written materials verbatim). In addition, a large majority of school directors (about 90 percent) reported that they believe the training is improving classroom instruction over time.

The program's theory of change assumed that changes in teaching practices would occur over longer periods of time, so in some ways the pattern of results in the interim evaluation report aligned well with the program's expectations. At the time of the interim report, it was not possible to assess whether changes in teaching behavior would take place in the longer term (one or two years after the end of the training sequence), as teachers develop new practices in the classroom. These longer-term changes are the focus of this chapter, which incorporates a new round of survey data from trained teachers and school directors that the study collected in fall 2019. Importantly, the data collection period of this evaluation did not align with the amount of time program implementers expected would be needed to observe changes in student learning outcomes. As a result, this study cannot say whether any of the changes in teacher outcomes discussed below ultimately led to improvements in student learning.

This chapter has two sections. The first section examines the potential impact of a recent policy implemented by the government of Georgia that provided an incentive for teachers to retire. Because the policy was adopted before the evaluation's final survey round, we sought to assess how the wave of retirements has affected the sample of teachers included in the TEE final analyses. The second section presents data the evaluation has collected on trends in teachers' knowledge and use of TEE-supported teaching practices. By examining trends in teaching practices up to two years after the training sequence ended, the analysis sheds light on whether trainees' knowledge of TEE-supported teaching practices remained high and whether teachers' use of these practices improved over time.

A. Accounting for teacher retirement incentives in 2019

In May 2019, the Ministry of Education announced a new retirement incentive program: teachers who had reached full retirement age were given a generous retirement incentive (valued at approximately two years of salary), and any eligible teachers who refused the offer would be required to pass a professional-competency exam to continue with their public school careers. The goal of the policy was to encourage older teachers who might be less willing to undergo new types of professional development to retire and make space for younger teachers to enter the classroom.

As a result of the program to encourage retirement, a substantial portion of the TEE evaluation sample (12 percent) retired and therefore did not participate in the September 2019 survey round. This sample attrition matters for the evaluation design, because the voluntary retirements may have led to systematic changes in the composition of the study sample in ways that are associated with the study's outcomes of interest (teaching practices). This could introduce bias into our final analyses. For example, if retirees were less (or more) likely to use a key teaching practice than teachers who did not retire, their absence could increase (or decrease) average reported use of the practice in the data simply due to the change in the sample of teachers surveyed.

The 2019 retirees were predominantly longstanding teachers who had not qualified for more senior professional classifications. To explore the potential magnitude and direction of this bias, we examined differences in demographics, teaching experience, and a selection of previously measured teaching practices between teachers who were interviewed in 2019 and teachers who retired in 2019 (Table IV.1). On average, the retirees were 20 years older (and had 20 years of additional teaching experience) compared to the non-retirees who remained in the teaching workforce and completed surveys in fall 2019. Despite the large difference in experience, nearly all retirees (99 percent) were classified as practitioner teachers (meaning they had not passed Georgia's teacher certification exam), compared to two-thirds of teachers interviewed in 2019. On average, the retirees taught two fewer hours each week and were less likely to teach elementary or lower-secondary grades (grades 1–6 and 7–9, respectively). Retirees were also more likely to teach science (and less likely to teach English), so we expect attrition to be more of an issue for final analyses of science and English teachers than for geography and math, where the subject-specific differences are smaller.

The 2019 retirees had substantially less exposure to TEE training. Retirees were 20 percentage points less likely to attend all TEE training modules, 27 percentage points less likely to attend all three core training modules, and 14 percentage points less likely to attend a single training module. Among teachers who had attended TEE training sessions, the retirees were also less likely to be "satisfied" or "very satisfied" with the training experience. Perhaps as a result, the retirees were less likely to use many of the teaching practices taught in the TEE training. For example, the retirees were 28 percentage points less likely to say they use informal assessments daily. In addition, the retirees were less likely to say they regularly ask open-ended questions, work with struggling students on a separate basis, or discuss inclusion of students with different ethnicities, religions, or sexual identities.

In short, the 2019 wave of retirements appears to have produced a teaching workforce that is younger, more likely to have completed the TEE training sequence, and more likely to use TEE-supported teaching practices. As a result, it is necessary to account for changes in the teaching workforce to avoid biasing the analysis of post-training trends in teaching practices. For the final report, we have elected to (1) present trend analyses only for the subsample of teachers who remained in the teaching workforce for all three data collection rounds (fall 2017, fall 2018, and fall 2019) and (2) omit the teachers who retired in 2019. Examining trends for a consistent sample will focus the descriptive results on changes that are more likely to be associated directly with the TEE program. This methodological decision should not trivialize or dismiss the importance of the government's retirement policy in any way. Indeed, it appears that the retirement incentives may have helped to achieve some of the same objectives as the TEE training initiative. That said, in our view it is more accurate to examine the effects of the retirement policy on a separate basis from the outcomes of the Compact's TEE training intervention.

Table IV.1. Teachers who retired in 2019 were less likely to attend TEE training sessions and report using TEE-supported practices

	Retired in 2019		Did not retire in 2019		Mean difference
	N	Mean	N	Mean	
Practitioner teacher	126	0.99	957	0.68	0.32***
Assigned to first TEE cohort					
Among all teachers	126	0.68	957	0.75	-0.07^
Among practitioner teachers	125	0.68	647	0.67	0.01
Male	129	0.14	970	0.11	0.03
Age (years)	129	69.0	970	49.1	19.8^
Teaching experience (years)	129	42.5	970	22.7	19.7^
Classroom hours per week	129	13.7	970	15.7	-2.0**
Grade levels taught					
Elementary (1–6)	129	0.41	970	0.59	-0.18**
Lower secondary (7–9)	129	0.86	970	0.92	-0.06*
Upper secondary (10–12)	129	0.84	970	0.83	0.01
Subjects taught					
Math	129	0.36	970	0.29	0.07^
Science	129	0.47	970	0.34	0.13**
Geography	129	0.13	970	0.14	0.00
English	129	0.07	970	0.27	-0.20**
Attended TEE training					
All training modules	104	0.59	902	0.79	-0.20**
All core training modules	106	0.58	903	0.85	-0.27**
Any training module	106	0.81	903	0.95	-0.14**
Satisfaction with TEE training					
Satisfied with training	94	0.88	894	0.95	-0.07**
Very satisfied with training	94	0.28	894	0.40	-0.12*
Teaching practices					
Ask open-ended questions: Every day?	129	0.38	970	0.48	-0.10*
Work with struggling students: Every day?	129	0.11	970	0.19	-0.08*
Use informal tests to assess learning: Every day?	129	0.20	969	0.48	-0.28**
Discuss inclusion of ethnicities/religions/sexual identities: Every month?	129	0.24	970	0.34	-0.10*
Discuss inclusion of special needs: Every month?	128	0.37	969	0.50	-0.14**

Source: Millennium Challenge Corporation Georgia Training Educators for Excellence Teacher Surveys (2017, 2018).

^/*/** indicates that differences are significant at the 10/5/1 percent levels.

B. Post-training trends in TEE-supported teaching practices

The final round of data collection for the TEE evaluation was designed to examine longer-term trends in teacher outcomes following TEE training. The analyses test whether teachers were able to retain or increase their knowledge of TEE-supported teaching practices during the follow-up period and to examine trends in how teachers report applying these practices in their classroom teaching.

The analysis revealed the following high-level findings:

- Two years after the TEE training sequence, nearly all teachers continued to report that they are "confident" or "very confident" in having enough knowledge to apply the student-centered instruction practices that were the focus of the TEE activity.
- During this period, the pattern of changes in teachers' self-reported instruction practices was mixed. Among the first cohort of trainees (the group we can observe for two post-training years), use of TEE-supported practices in science instruction increased over time, but there were only modest changes in the use of other student-centered teaching practices. This initial training cohort prioritized more highly qualified teachers who had passed Georgia's teacher certification exam.
- However, among the second cohort of trainees (who we observed for one year after the training sequence ended), there were large improvements in the use of teaching practices related to students' critical thinking and collaboration, such as asking open-ended questions and having students present their work. Unlike the first cohort, the second cohort consisted largely of practitioner teachers (meaning they had not passed Georgia's teacher certification exam), and they initially reported using student-centered practices less often than Cohort 1 teachers shortly after the training sequence ended. As of fall 2019, these teachers in the second cohort appear to have caught up with more qualified Cohort 1 teachers in their use of certain TEE-supported instructional practices.

The remainder of this chapter discusses the results in greater detail. The analysis presented here relies on survey data collected for the TEE evaluation at three points in time: September 2017 (one month after completion of the full training sequence for Cohort 1), September 2018 (one month after completion of the full training sequence for Cohort 2), and September 2019 (to measure longer-term post-training outcomes for both cohorts). Because of the phased implementation of the TEE training, we can observe teacher outcomes two years after training was complete for teachers in Cohort 1. For the teachers in Cohort 2, we observe teaching outcomes after the first full year after the training sequence ended.

To ensure that the samples are comparable across time (especially considering the effect of the retirements discussed in the prior section of this chapter), all the trend analyses in this section are restricted to teachers who remained in the teaching workforce and responded to the TEE survey in each of the study's three data collection rounds. For example, we restricted the analysis of changes in each teaching practice between 2017 and 2019 to Cohort 1 teachers who completed the training sequence in fall 2017, remained in the teaching workforce, and completed the evaluation survey in both fall 2018 and fall 2019.

Among Cohort 1 teachers, self-reported knowledge of TEE-supported practices remained high two years after the training. In the first month after the TEE training sequence ended, nearly all Cohort 1 teachers reported that they were either "confident" or "very confident" in their knowledge of various TEE-supported teaching practices (Table IV.2). This remained true two years later, and demonstrates that overall confidence levels remained high in both time periods. However, immediately after the training sequence ended, 20 to 35 percent of the trainees reported they were "very confident" in their knowledge of these practices. Two years later, significantly fewer teachers reported they were "very confident" in their knowledge of each of the practices, with the declines for each ranging from 4 to 14 percentage points. These changes in self-reported knowledge do not necessarily imply that teachers are using these practices less often. Indeed, some teachers may have felt very confident in their knowledge initially after training but learned that some practices were more challenging than anticipated precisely because they had begun applying new practices in the classroom (which was the intended outcome of the training sequence). We examine trends in classroom practices next.

Table IV.2. Self-reported knowledge of teaching practices, two years after training

	Confident or very confident in knowledge			Very confident in knowledge		
	Post-training mean		Change between time periods	Post-training mean		Change between time periods
	One month after	Two years after		One month after	Two years after	
	2017	2019	2017–2019	2017	2019	2017–2019
Practices related to critical thinking, motivation, and collaboration						
Teaching to motivate and encourage	0.96	0.98	0.02*	0.33	0.19	-0.14**
Teaching to build self-confidence	0.95	0.97	0.02*	0.29	0.20	-0.09**
Teaching to build higher-order thinking	0.96	0.97	0.01	0.32	0.18	-0.14**
Promoting cooperation through group work	0.96	0.98	0.02*	0.28	0.15	-0.13**
Practices related to tailoring learning to student needs						
Creating a lesson plan with different tasks	0.92	0.93	0.02	0.21	0.17	-0.04*
Practices related to assessing student learning						
Conceptualizing measurable learning objectives	0.93	0.95	0.02	0.27	0.18	-0.08**
Using formative assessments during lessons	0.97	0.98	0.01	0.31	0.20	-0.11**
Including formative assessments in lesson plans	0.95	0.96	0.01	0.24	0.17	-0.07**
Practices related to inclusion						
Creating equitable learning environment for girls	0.93	0.96	0.03*	0.24	0.15	-0.08**
Creating equitable learning environment for students with special needs	0.85	0.91	0.06**	0.22	0.16	-0.06**
Creating unbiased learning environment	0.97	0.98	0.01	0.32	0.19	-0.13**
Practices related to ICT use						
Using ICT in instruction	0.92	0.95	0.03**	0.35	0.21	-0.14**

Source: Training Educators for Excellence Evaluation Teacher Surveys (2017, 2019). N = 712.

^/*/** indicates that differences are significant at the 10/5/1 percent levels

ICT= information communication and technology

In our interim report, we found little evidence of changes in teaching practices in the first month after the end of the TEE training sequence. However, changes in teaching practices might require more time to develop for several reasons. With additional time, trained teachers might have started (1) testing ways to apply the knowledge gained in the TEE trainings; (2) using the most helpful practices on a consistent basis; and (3) experiencing greater spillover benefits among teachers beginning in the 2018–2019 school year, after all the teachers in the study schools completed the training sequence. In addition, Cohort 1 teachers were given an opportunity to make up for any training modules that they missed and sit in on training events attended by Cohort 2 teachers. (As a result, the percentage of Cohort 1 teachers who attended any TEE training modules increased from 64 to 82 percent after the second round of training.) This additional exposure to training among Cohort 1 teachers may have affected their teaching practices in later years.

Although Cohort 1 teachers reported few long-term changes in their use of student-centered teaching practices, they did report improvements in teaching practices related to science instruction. As shown in Table IV.3, the changes in student-centered teaching practices were modest in size, ranging from a decline of 5 percentage points (for example, in the use of daily lesson plans designed to achieve specific learning goals) to an increase of 4 percentage points (for making daily changes in instruction based on testing). However, there was a notably larger improvement in science-related teaching practices: during the two years after the training sequence ended, monthly use of lab experiments increased by 7 percentage points and monthly practice of hypothesis testing increased by 8 percentage points.

The practitioner teachers in Cohort 1 show a different pattern of long-term changes in teaching practices, compared with more qualified teachers. Although Cohort 1 teachers were more likely to meet the government's professional certification criteria (these more qualified staff are classified in Georgia as "lead," "mentor," or "senior" teachers), a meaningful portion of the first cohort consisted of practitioner teachers whose outcomes differed in notable ways. As shown in Table IV.4, compared with practitioner teachers, we find slightly larger declines among more qualified teachers for collaborative group work, students presenting work, preparing lesson plans to achieve specific learning goals, and using formal tests to assess learning. The improvements in science-specific practices also appear to be largely driven by changes among practitioner teachers, who increased their use of laboratory experiments and hypothesis formation and testing by 9 percentage points (compared to 2 and 5 percentage points, respectively, among more qualified teachers). On the other hand, more qualified teachers outpaced practitioner teachers in improving their frequency of adjusting instruction in response to tests and using information technology for classroom instruction.

Table IV.3. Changes in reported Cohort 1 teaching practices between one month and two years after training

	Post-training mean		Change between time periods
	One month after	Two years after	
Student-centered teaching practices	2017	2019	2017–2019
Practices related to critical thinking, motivation, and collaboration			
Ask open-ended questions: Every day?	0.51	0.53	0.02
Collaborative group work: At least three times per week?	0.40	0.36	-0.04^
Students present work: At least three times per week?	0.44	0.43	-0.01
Students work independently: Every day?	0.52	0.47	-0.05*
Practices related to tailoring learning to student needs			
Lesson plans include differentiated activities: Every day?	0.11	0.13	0.03
Work with struggling students: Every day?	0.21	0.16	-0.05**
Practices related to assessing student learning			
Prepare lesson plans to achieve specific learning goals: Every day?	0.39	0.34	-0.05*
Use formal tests to assess learning: At least once per week?	0.63	0.58	-0.05*
Use informal tests to assess learning: Every day?	0.48	0.47	-0.02
Change instruction in response to tests: Every day?	0.20	0.24	0.04*
Practices related to inclusion			
Discuss inclusion of ethnicities/religions/sexual identities: Every month?	0.35	0.33	-0.02
Discuss inclusion of girls: Every month?	0.46	0.45	-0.01
Discuss inclusion of students with special needs: Every month?	0.49	0.50	0.01
Practices related to ICT use			
Use ICT in instruction: Every week?	0.50	0.52	0.03
Practices related to teaching science courses			
Students conduct laboratory experiments: At least once per month?	0.57	0.64	0.07*
Students practice making or testing hypotheses: At least once per month?	0.74	0.83	0.08*

Source: Training Educators for Excellence Evaluation Teacher Surveys (2017, 2019). N = 717 for general practices; N = 256 for science practices.

^/*/** indicates that differences are significant at the 10/5/1 percent levels.

ICT= information communication and technology

Table IV.4. Changes in reported Cohort 1 teaching practices between one month and two years after training, by teacher qualification level

Teaching practices	Change over two years, by teacher qualification level	
	Senior, lead, or mentor	Practitioner
	2017–2019	2017–2019
Practices related to critical thinking, motivation, and collaboration		
Ask open-ended questions: Every day?	0.02	0.02
Collaborative group work: At least three times per week?	-0.08*	-0.01
Students present work: At least three times per week?	-0.08*	0.04
Students work independently: Every day?	-0.06	-0.04
Practices related to tailoring learning to student needs		
Lesson plans include differentiated activities: Every day?	0.02	0.03
Work with struggling students: Every day?	-0.05	-0.06*
Practices related to assessing student learning		
Prepare lesson plans to achieve specific learning goals: Every day?	-0.06^	-0.04
Use formal tests to assess learning: At least once per week?	-0.07^	-0.04
Use informal tests to assess learning: Every day?	-0.02	-0.01
Change instruction in response to tests: Every day?	0.10**	0.01
Practices related to inclusion		
Discuss inclusion of ethnicities/religions/sexual identities: Every month?	0.02	-0.05
Discuss inclusion of girls: Every month?	0.03	-0.04
Discuss inclusion of students with special needs: Every month?	0.02	0.01
Practices related to ICT use		
Use ICT in instruction: Every week?	0.07^	0.00
Practices related to teaching science courses		
Students conduct laboratory experiments: At least once per month?	0.02	0.09*
Students practice making or testing hypotheses: At least once per month?	0.05	0.09*

Source: Training Educators for Excellence Evaluation Teacher Surveys (2017, 2019). N = 287 for senior, lead, or mentor teachers; N = 432 for practitioner teachers.

^/*/** indicates that difference between the 2017 mean and 2019 mean for a given group of teachers is significant at the 10/5/1 percent levels

ICT= information communication and technology

While the pattern of longer-term changes for Cohort 1 teachers is mixed (and varies somewhat with teachers' seniority), among Cohort 2 teachers there was a strong pattern of improvements over time in practices related to critical thinking, motivation, and collaboration. Nearly all teachers in Cohort 2 are practitioner teachers—in the first month after the training sequence ended, these teachers were using practices related to critical thinking, motivation, and collaboration significantly less often than the more qualified teachers in Cohort 1. However, Cohort 2 teachers reported statistically significant improvements during the first post-training year in their use of open-ended questions (an increase of 9 percentage points), collaborative group work (an increase of 10 percentage points), and having students present their work (an increase of 22 percentage points), as shown in Table IV.5.[17]

Table IV.5. Changes in reported Cohort 2 teaching practices one year after training

Student-centered teaching practices	Post-training mean: One month after 2018	Post-training mean: One year after 2019	Change between time periods 2018–2019
Practices related to critical thinking, motivation, and collaboration			
Ask open-ended questions: Every day?	0.40	0.49	0.09*
Collaborative group work: At least three times per week?	0.29	0.39	0.10*
Students present work: At least three times per week?	0.28	0.50	0.22**
Students work independently: Every day?	0.45	0.49	0.05
Practices related to tailoring learning to student needs			
Lesson plans include differentiated activities: Every day?	0.11	0.14	0.03
Work with struggling students: Every day?	0.21	0.20	-0.02
Practices related to assessing student learning			
Prepare lesson plans to achieve specific learning goals: Every day?	0.43	0.34	-0.09*
Use formal tests to assess learning: At least once per week?	0.66	0.64	-0.02
Use informal tests to assess learning: Every day?	0.50	0.48	-0.02
Change instruction in response to tests: Every day?	0.20	0.23	0.03
Practices related to inclusion			
Discuss inclusion of ethnicities/religions/sexual identities: Every month?	0.32	0.35	0.02
Discuss inclusion of girls: Every month?	0.45	0.43	-0.02
Discuss inclusion of students with special needs: Every month?	0.49	0.45	-0.04
Practices related to ICT use			
Use ICT in instruction: Every week?	0.48	0.54	0.06^
Practices related to teaching science courses			
Students conduct laboratory experiments: At least once per month?	0.63	0.67	0.04
Students practice making or testing hypotheses: At least once per month?	0.84	0.94	0.10^

Source: Training Educators for Excellence Evaluation Teacher Surveys (2018, 2019). N = 220.

^/*/** indicates that differences are significant at the 10/5/1 percent levels

ICT= information communication and technology

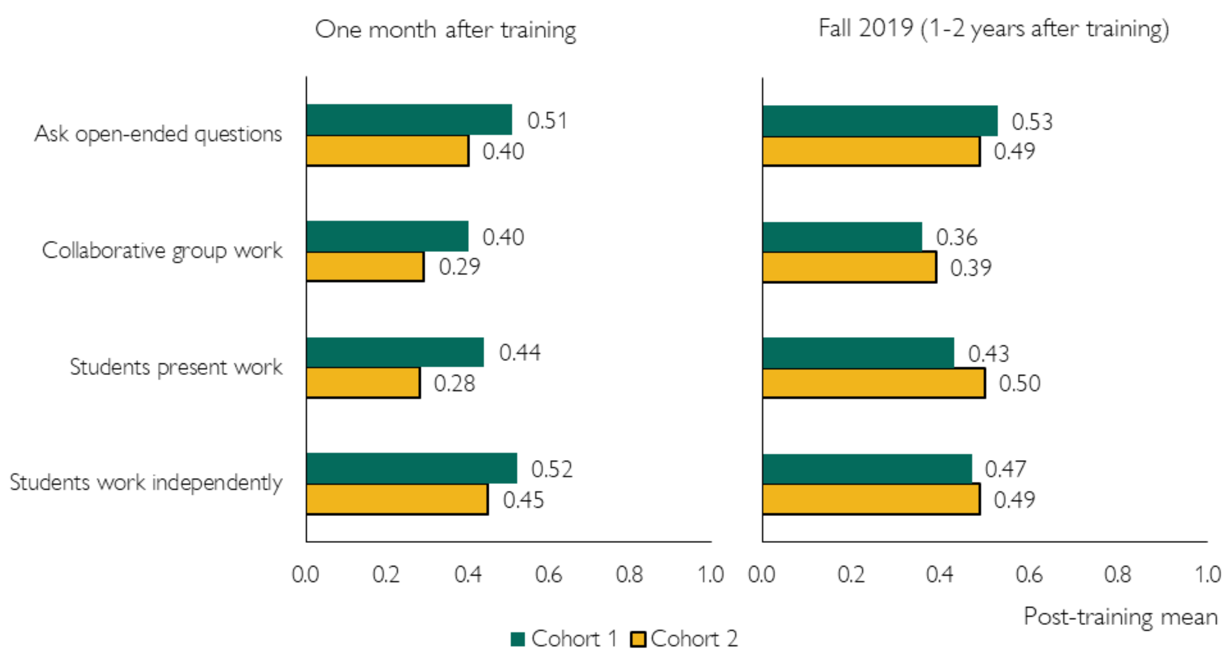
In other domains the changes for Cohort 2 were smaller, with a mixed pattern of modest increases and decreases. The only statistically significant decline (9 percentage points) was in the percentage of Cohort 2 teachers who prepare lesson plans to meet specific learning goals daily. Cohort 2 teachers reported a modest improvement in monthly use of science-related practices (by 4 percentage points in lab use and 10

[17] The pattern of improvements over time in practices related to critical thinking, motivation, and collaboration are calculated by comparing the means of the reported use of teaching practices one month after training (presented in the first column in Table IV.3 [Cohort 1]) with the means in the fall of 2019 (presented in the first column of Table IV.5 [Cohort 2]).

percentage points in hypothesis testing), but neither change was statistically significant. Similarly, none of the other measures of TEE-related practices showed a change that was statistically significant.

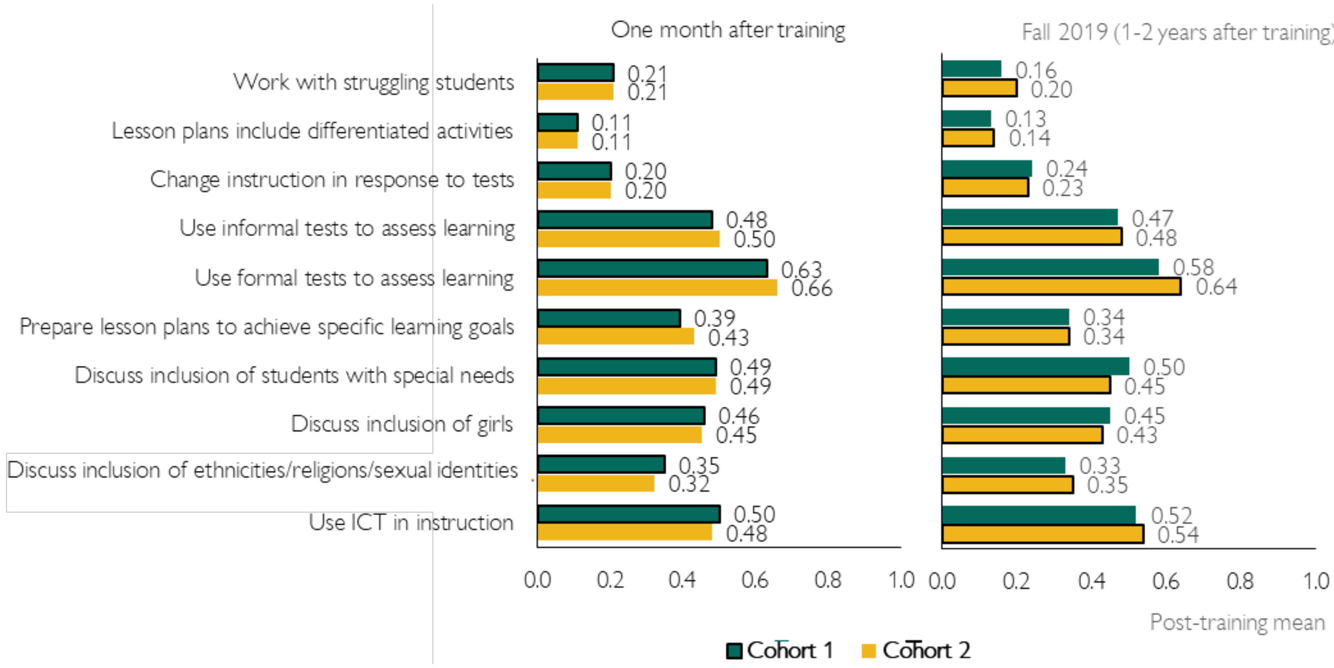
The improvements observed among Cohort 2 teachers effectively closed the gap between them and the more qualified teachers in Cohort 1 for practices related to critical thinking, collaboration, and motivation. Immediately after each cohort completed training, Cohort 1 teachers consistently outperformed Cohort 2 teachers in this domain (Figure IV.1). However, the longer-term analysis revealed that Cohort 1 teachers remained relatively stable in how often they used these practices over the following two years, whereas Cohort 2 teachers showed large improvements over the first 12 months after completing the training sequence. Descriptively, comparing the post-training means in the two cohorts (that is, subtracting the fall 2017 means for Cohort 1 [Table IV.3] from the fall 2018 means for Cohort 2 [Table IV.5]) shows that, immediately after training, Cohort 1 outperformed Cohort 2 across all of the practices in the domain by a magnitude of 7 to 16 percentage points. By the endline survey in fall 2019, however, Cohort 2 teachers were only underperforming Cohort 1 teachers in their use of open-ended questions (by a magnitude of 4 percentage points), and they were slightly outperforming Cohort 1 teachers on the other three practices (by a magnitude of 2 to 7 percentage points).

Figure IV.1. Comparison of reported practices related to critical thinking, collaboration, and motivation between one month and two years after completing training



One year after training, Cohort 2 teachers were also keeping up with teachers in Cohort 1 in the other TEE program domains of tailored learning, assessments, inclusion, and information communication technology (ICT) use. Improvements over time in the 12 months after training were less visible in these other domains. However, in the final survey round, Cohort 2 (which largely consisted of practitioner teachers) reported that they were using practices in these areas at a very similar rate to the teachers in Cohort 1 (who were more qualified, on average). Thus, in these domains, Cohort 2 teachers may have already caught up with the teachers in Cohort 1 shortly after the training sequence came to an end. For example, in the domain of tailored learning practices, Cohort 2 teachers slightly outperformed Cohort 1 teachers by 1 to 4 percentage points in fall 2019 (Figure IV.2). Similarly, as of fall 2019, teachers in both cohorts reported very similar usage rates for practices related to the assessment domain (within 6 percentage points), practices related to the inclusion domain (within 5 percentage points), and with respect to the use of information technology (within 2 percentage points).

Figure IV.2. Comparison of reported practices between one month and two years after training



C. Interpreting the TEE study's final results

The final evaluation of the TEE activities provides new information about how the effects of these training activities have changed over time. Our first set of findings showed that Georgia's incentivized teacher retirement program has changed the composition of teachers in the country (and the TEE evaluation sample): we found that retirees were more likely to be older, practitioner-level teachers who appeared to be less enthusiastic about the TEE training and TEE-supported practices. These selective changes in the composition of the TEE evaluation sample would likely skew TEE final results if we used the full sample, so we limited the main trends analysis to a consistent sample of teachers who participated in each survey round and chose not to retire in 2019. That said, it is notable that the retirement program appears to have been an effective complement to the goals of the TEE activity: compared to the teachers who remained, the 12 percent of teachers who retired in 2019 were substantially less likely to have adopted the types of teaching practices that were encouraged during the Compact.

The second set of findings examined changes in knowledge and use of TEE-supported teaching practices in the post-training period. Teachers reported that their general knowledge of the training material remained high, albeit with some fading among teachers who had been most confident coming out of the training. For Cohort 1 teachers (who were more qualified on average than the teachers in Cohort 2), we found evidence of long-term improvements in the use of science-related practices (having students conduct laboratory experiments and practice testing hypotheses at least once a month), but relatively little evidence of improvements in other practices. For Cohort 2 (which largely consisted of less qualified practitioner teachers), we found evidence that they were catching up to Cohort 1 in the use of practices related to critical thinking, motivation, and collaboration, one year after the training sequence ended.

There are several potential explanations for the pattern of improvements observed among Cohort 2 teachers. One potential explanation for this pattern could be that the TEE activity focused on practices that practitioner teachers required more time to adopt. In the first month after the training sequence ended, practitioner teachers in Cohort 2 were using TEE-related practices significantly less often than the teachers in Cohort 1. Because practitioner teachers had yet to earn the higher qualification levels (and corresponding salary increases) available to senior, lead, and mentor teachers, teachers at the practitioner level may have been more motivated improve their instruction practices over an extended period of time. Alternatively, since more qualified teachers were already using these practices substantially in the first follow-up survey (immediately after training), there may have been a plateau effect among this group if they saw little reason to change practices beyond what they put in place immediately. Finally, the TEE activity may have found ways to improve the quality of post-training oversight and support for collaboration among teachers by applying lessons learned with the first cohort. If the activity provided greater post-training support to the Cohort 2 teachers, it may have helped them to demonstrate improving trends that were largely absent in Cohort 1.

This page has been left blank for double-sided copying.

V. Conclusion

This report has presented the final set of findings from the Georgia IGEQ Project evaluation using data collected up to two years after the end of the project's five-year implementation period. In this concluding chapter, we summarize how the findings in this report have contributed to answering the study's evaluation questions.

A. Final set of findings about the school rehabilitation activity

We first summarize the key findings related to each of the evaluation questions for the school rehabilitation activity covered in this final report. This activity successfully rehabilitated infrastructure in a total of 91 school buildings (88 of which were included in the evaluation sample), producing large improvements in the learning environment which were highly visible to and appreciated by students, teachers, and parents. However, the program also encountered challenges related to accessing and using upgraded science labs in these schools, and the longer-term effects of these investments on learning outcomes are ambiguous at this point. Key findings from the final report, organized by evaluation question, follow.

RQ1. What are the impacts of the ILEI activity on the school infrastructure environment, such as temperature, maintenance policy, and maintenance practice?

The school rehabilitation activity dramatically improved the quality of rehabilitated buildings. Findings from the evaluation's randomized control trial reveal that the activity had a consistent and large impacts on a wide range of infrastructure outcomes. Some of the key upgrades in rehabilitated schools included eliminating glaring and widespread problems with classroom walls, ceilings, and floors; installing electrical lighting systems in classrooms that had no functional lighting before; installing central heating systems that improved classroom temperatures and eliminated the serious air quality problems associated with wood-burning stoves; and upgrading sanitary facilities (with running water and flush toilets) and science labs (with lab benches and equipment for experiments and demonstrations) in highly visible ways that were noticed and appreciated by students and teachers. While school directors reported an increase in operating costs associated with these investments (particularly utility costs related to using a central heating system), there is little evidence these costs were unmanageable. In some cases, the heating costs in rehabilitated schools were offset by revenues from increased student enrollment (particularly in early grades), as more families chose to enroll children in rehabilitated school instead of other regional schools that did not receive the program.

RQ2. What are the impacts of the ILEI activity on teachers' behavior, such as attendance and time spent teaching?

Teachers reported that these infrastructure upgrades addressed multiple serious problems that had limited their ability to use classroom time effectively. In surveys and in-depth qualitative interviews, teachers consistently reported that infrastructure upgrades directly improved their ability to focus on instruction in the classroom. Functional electric lighting made it possible for students to read written materials more easily, access to indoor sanitary facilities helped students reach their classrooms in less time, and in winter teachers reported that central heating brought their classrooms to a more comfortable temperature (meaning students no longer needed to wear winter jackets during lessons) and addressed serious air quality problems related to poor ventilation and smoke from wood-burning stoves. While these changes did not affect teachers' overall attendance rates or the total number of minutes teachers spent working in the classroom each day, respondents consistently pointed out that the activity made it possible to focus more fully on instruction without the discomforts and distractions present in the building before rehabilitation took place.

However, teachers also reported difficulties with accessing and using upgraded science labs. Rehabilitated school buildings received a single upgraded science lab, which was designed to be shared across secondary grade-levels and with equipment spanning multiple subjects (including biology, chemistry, and physics). It is clear that these upgraded science labs are being used actively (for example, the rehabilitation improved the likelihood that students will see a science demonstration or experiment during the school year). That said, science teachers in rehabilitated schools consistently reported that they have not been able to use the labs as often as they had hoped to integrate the new facilities into their curriculum and daily lessons. Scheduling access to the labs is difficult due to the number of grade levels and subjects seeking to access the facility, and some teachers also reported that it is difficult find time to store and set up equipment for their subject when the lab has been arranged for a lesson in a different subject. In some schools, the amount of equipment storage needed in the labs also impeded airflow from the central heating system and caused the labs to be uncomfortably cold in winter. All these frustrations were further exacerbated by requirements to enforce social-distancing regulations during the COVID-19 pandemic: because the layout of these labs are organized around conducting small-group experiments at shared tables, in many schools it was not possible to access the labs at all during the 2020–2021 or 2021–2022 school years.

RQ3. What were the impacts of the ILEI activity on student outcomes such as attendance and dropout rates, time spent studying in and out of school, and learning outcomes?

As with teachers, students also reported that infrastructure upgrades addressed serious barriers to learning in the classroom. The program did not have an impact on student absenteeism or dropout rates. In the case of dropout patterns, these rates were already quite low before the program began (averaging less than 5 percent per year, across upper-secondary grades), suggesting there was only limited room for improvement. Consequently, there is little evidence that school rehabilitation changed the total number of days students spend at school in a given year. However, the activity substantially improved the quality of the time students spent on learning activities during the school day. In surveys and focus groups, data from students in rehabilitated schools revealed a striking pattern of improvements in views about their ability to focus on learning in the classroom. In schools that were not rehabilitated (the evaluation's control group), students pointed out that poor lighting, inadequate heating, harmful air quality, ceiling leaks, lack of indoor toilets, and the absence of science labs were all serious problems affecting their comfort and ability to focus on instruction. None of these issues remained in rehabilitated schools. In addition to enhancing student comfort and ability to focus, rehabilitation also directly reduced interruptions during the school day (for example, the need to pause lessons to refuel stoves and change rooms to air out classrooms when wood smoke became overwhelming). By eliminating these distractions, the program directly increased learning time.

However, the short-term effects of rehabilitation on learning outcomes remain ambiguous, in part due to the disruptions of the COVID-19 pandemic. After two years of access to rehabilitated schools, we did not find evidence that the activity had an impact on math, language, or science test scores. An important complicating factor in this analysis is that the school closures caused by the COVID-19 pandemic occurred approximately halfway through the follow-up period in this evaluation: some schools in the sample completed the study's two-year follow-up period before the pandemic began, whereas other schools experienced extended disruptions and school closures during the evaluation's follow-up period. While this study was not designed to carrying out precise subgroup analyses testing for differences between these two groups schools, the data suggest that learning outcomes may have differed dramatically during the pandemic. Particularly in language and math, rehabilitation appears to have had a positive effect on learning outcomes before the pandemic—but these effects were absent for schools assessed after the pandemic. If the pre-pandemic impacts of rehabilitation had held for the entire sample, impacts of that magnitude would have represented a meaningful amount of learning growth. For the language exam, the pre-pandemic impact of 0.13 standard deviations is equivalent to about seven months of learning for upper-secondary students. Similarly, pre-pandemic impact on math scores (0.10 standard deviations) is equivalent to five months of learning. On the other hand, in science, the pre-pandemic effect of rehabilitation was close to zero—an outcome which might be explained by the barriers to science-lab access and use discussed above.

RQ4. What are the long-term impacts of the ILEI activity?

Rehabilitated schools have been able to maintain improved infrastructure over time, and there is some evidence of longer-run improvements in learning outcomes as well. Follow-up data collection activities conducted three to five years after rehabilitation was completed show that these schools have been able to maintain infrastructure improvements over time: rehabilitation continued to produce a dramatic improvement in overall building quality, heating systems and heating-system usage, air quality, sanitation facilities, and perceptions of school safety and the perceived quality of the learning environment among students and teachers. Interestingly, this data collection round also revealed suggestive evidence that rehabilitation may produce more positive learning benefits beyond the second follow-up year. In all three subjects (language, math, and science), impacts on learning outcomes were negative or close to zero for schools in their second follow-up year, but impacts became positive (and larger in magnitude) in schools that were in their third, fourth, or fifth follow-up year. Although these results are only suggestive, longer-term impacts of this magnitude (0.12 standard deviations in language, 0.11 standard deviations in math, and 0.11 standard deviations in science) would represent an educationally meaningful boost in student achievement. For example, an effect size of 0.10 standard deviations would be roughly equivalent to five months of learning in language and seven months of learning in math, for students in grade 10.

B. Final set of findings about the TEE activity

In this section, we summarize the key findings for the final evaluation of the TEE activity. This study's interim evaluation report provided an initial set of results for each of the TEE evaluation's research questions and showed that the TEE activity succeeded in implementing the program on a nationwide scale. Simply implementing a training intervention on this scale was a remarkable achievement: the training sequence was offered to Georgia's entire population of school directors (about 2,000) and all of Georgia's upper-grade teachers in the subjects of science, mathematics, English, and geography (about 18,000 teachers in total). In terms of the training's potential effects, in the interim analysis we also found a consistent pattern of improvements in teachers' self-reported knowledge of student-centered instruction strategies in the initial period after training, approximately one month after finishing the one-year training sequence. However, outside of professional development activities (where we found a stronger pattern of improvements), the interim analysis did not reveal consistent evidence of short-term changes in teachers' classroom practices.

For the TEE activity, it is important to remember that the program logic did not assume that teaching practices would change in the immediate aftermath of the training sequence. Instead, the program was designed to produce rapid improvements in teachers' knowledge and their professional development resources (through the use of teacher study groups and other professional networks), which would in turn produce changes in their teaching practices and ultimately improve students' learning outcomes over longer periods of time. To examine whether this pattern occurred, this final evaluation report included a longer-term follow-up analysis of teachers' and school directors' practices up to three years after the training sequence was completed. The key findings from the final analysis are as follows.

The Georgian government's 2019 incentivized retirement policy produced a teaching workforce that is younger, more likely to have completed the TEE training sequence, and more likely to be using TEE-supported teaching practices. In May 2019, the Ministry of Education announced a new retirement incentive program: teachers who had reached full retirement age were given a generous retirement incentive, and any eligible teachers who refused the offer would be required to pass a professional-competency exam to continue with their public-school careers. The goal of the policy was to encourage older teachers who might be less willing to undergo new types of professional development to retire and make space for younger teachers to enter the classroom. Our analyses show that the policy appears to have worked exactly as designed and directly supported the goals of the TEE activity. These incentivized retirees comprised 12 percent of the evaluation's sample and were (on average) 20 years older than non-retirees, 27 percentage points less likely to have completed the core TEE training modules, and substantially less likely report using TEE-related instructional practices such carrying out daily informal assessments, asking open-ended questions, working with struggling students on a separate basis, or discussing inclusion of students with different ethnicities, religions, or sexual identities.

Two years after the TEE training sequence, nearly all teachers continued to report that they are "confident" or "very confident" in having enough knowledge to apply the student-centered instruction practices that were the focus of the TEE activity. As with the interim evaluation, the final survey round revealed that teachers remained highly confident that they had acquired enough knowledge to apply the types of teaching practices that were the focus of the TEE training sequence. There is very little evidence that the knowledge gains reported by teachers shortly after the training sequence ended have meaningfully faded over time.

Among the first cohort of trainees (which largely consisted of highly qualified teachers who had already passed Georgia's teacher certification exam), reported use of TEE-supported practices in science instruction increased over time, but there were only modest changes in reported use of other student-centered teaching practices. Over a two-year period after the training sequence ended, among these more qualified teachers there were only modest changes in the reported use of student-centered teaching practices, ranging from a decline of 5 percentage points (for example, in the use of daily lesson plans designed to achieve specific learning goals) to an increase of 4 percentage points (for making daily changes in instruction based on testing). However, there was a notably larger improvement in science-related teaching practices: during the two years after the training sequence ended, monthly use of lab experiments increased by 7 percentage points and monthly practice of hypothesis testing increased by 8 percentage points.

However, among the second cohort of trainees (which consisted of less qualified practitioner teachers) there were large improvements in the use of teaching practices related to students' critical thinking and collaboration, such as asking open-ended questions and having students present their work. Interestingly, the second cohort consisted largely of practitioner teachers who had not passed

Georgia's teacher certification exam prior to the training. Shortly after the training sequence ended, this group of less qualified teachers initially reported using student-centered practices less often than the more highly qualified teachers in Cohort 1. As of fall 2019, Cohort 2 teachers (who were generally less qualified) appear to have caught up with Cohort 1 teachers (who were more qualified on average) in their use of certain TEE-supported instructional practices. For example, one year after the training sequence ended, Cohort 2 teachers reported statistically significant improvements in their use of open-ended questions (an increase of 9 percentage points), collaborative group work (an increase of 10 percentage points), and having students present their work (an increase of 22 percentage points). Immediately after training, teachers in the first cohort outperformed the practitioner teachers in Cohort 2 by a magnitude of 7 to 16 percentage points across these practices. By the endline survey in fall 2019, Cohort 2 teachers were only underperforming Cohort 1 teachers in their use of open-ended questions (by a magnitude of 4 percentage points), and they were slightly outperforming Cohort 1 teachers on the other three practices (by a magnitude of 2 to 7 percentage points).

C. Lessons from the final evaluation report

This final evaluation report from the evaluation of the IGEQ Project has important implications for the design and implementation of future evaluations of education investments and programs. The evaluation included two separate studies, each of which had a distinct intervention and theory of change. A cross-cutting lesson from both studies is that implementers and evaluators need to consider the expected *timing* of each step in an activity's theory of change. This includes accounting for the timing of programmatic inputs and outputs (completion of school rehabilitation or training sequences, across various cohorts of beneficiaries) alongside the expected timing of changes in short- and medium-term outcomes (such as improvements in the learning environment or changes in teaching practices) and longer-term outcomes (such as improved student learning).

In particular, the evaluation of these two interventions demonstrated that it is profoundly important to measure outcomes on a schedule that aligns with the timeline of each activity's expected theory of change. For the school rehabilitation study, for example, it would have been misleading to measure the impacts of school rehabilitation without adjusting for the periods of time when school buildings were closed due to the COVID-19 pandemic. Similarly, in the school rehabilitation study the general pattern of impacts on student learning outcomes appears to be quite different (and more favorable) after students had access to improved school infrastructure for more than two years. In the case of the TEE study, the final evaluation report also showed that continuing to track teaching practices for a year or longer can reveal substantial patterns of post-training changes (particularly among less qualified teachers who had the most room for improvement in their teaching practices) that were not visible at an earlier point in time.

Ultimately, the school rehabilitation activity and the TEE activity both produced substantial and important changes across some (but not all) of the outcomes that were the central focus of the IGEQ Project's designers and implementers. But the timing of these changes varied widely across outcomes. In light of these patterns, researchers and policymakers should remain open to the possibility that there could continue to be additional changes beyond the time-period of this evaluation. If teachers continue to gain mastery of enhanced teaching practices in the future, in schools where improved infrastructure continues to supply a dramatically improved learning environment, there is a real possibility that student learning outcomes could show further improvements over time. The ultimate effects of these two investments in Georgia's education system remain to be seen.

This page has been left blank for double-sided copying.

References

Bagby, Emilie, Anca Dumitrescu, Cara Orfield, and Matt Sloan. "Niger IMAGINE Long-Term Evaluation." Washington, DC: Mathematica Policy Research, October 24, 2014.

Bagby, Emilie, Kristine Bos, Anca Dumitrescu, Nicholas Ingwersen, and Matt Sloan. "Niger NECS Impact Evaluation Report." Washington, DC: Mathematica Policy Research, July 17, 2017.

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.

Bruns, Barbara and Javier Luque. *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*. Washington, DC: World Bank, 2015.

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.

Chetty, Raj, John N. Friedman, and Johnah E. Rockoff. "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." National Bureau of Economic Research Working Paper Series, working paper no. 17699. Cambridge, MA: National Bureau of Economic Research, December 2011.

Davis, Mikal, Nick Ingwersen, Harounan Kazianga, Leigh Linden, Arif Mamun, Ali Protik, and Matt Sloan. "Ten-Year Impacts of Burkina Faso's BRIGHT Program." Washington, DC: Mathematica Policy Research, August 29, 2016.

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 13, 2011.

Dunteman, George H. *Principal Components Analysis*. Newbury Park, California: SAGE Publications, 1989.

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.

Evans, David K., and Anna Popova. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." World Bank Policy Research Paper 7203. Washington, DC: World Bank Group, February 2015.

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.

Gilraine, Michael, and Angela Zhang. "Air Pollution and Student Performance in the U.S." Rochester, NY: Social Science Research Network, February 2022.

Hair, Joseph F., Ralph E. Anderson, Ronald L. Tatham, and William C. Black. *Multivariate Data Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall, 1998.

Hanushek, Eric. "The Economic Value of Higher Teacher Quality." National Center for the Analysis of Longitudinal Data in Education Research Working Paper 56. Washington, DC: Urban Institute, December 2010.

He, F., L. Linden, and M. Macleod. "A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial." Working paper. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab, Massachusetts Institute of Technology, 2009.

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.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*, vol. 2, no. 3, 2008, pp. 172–177.

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.

Kazianga, Harounan, Dan Levy, Leigh L. Linden, and Matt Sloan. "The Effects of 'Girl Friendly' Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso." *American Economic Journal: Applied Economics*, vol. 5, no. 3, 2013, pp. 41–62.

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." Final report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, December 2013.

Nichols-Barrer, Ira, Nicholas Ingwersen, Elena Moroz, and Matt Sloan. "Baseline Report for the Georgia Improving General Education Quality Project's School Rehabilitation Activity." Final report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, December 2017.

Nichols-Barrer, Ira, Nicholas Ingwersen, Elena Moroz, and Matt Sloan. "Evaluation Interim Report for the Georgia Improving General Education Quality Project's School Rehabilitation and Training Activities." Final report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, October 2019.

Palacios, Juan, Piet Eichholtz, Nils Kok, and Nicholas Duran. "Indoor Air Quality and Learning: Evidence from A Large Field Study in Primary Schools." Rochester, NY: Social Science Research Network, December 2022.

Popova, A., D.K. Evans, and V. Arancibia. "Training Teachers on the Job: What Works and How to Measure It." World Bank Group: Policy Research Working Paper. Washington, DC: World Bank, 2016.

Sailors, H. "The Effects of First- and Second-Language Instruction in Rural South African Schools." *Bilingual Research Journal*, vol. 33, no. 1, 2010, pp. 21–41.

Stevens, James. *Applied Multivariate Statistics for the Social Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.

WHO Regional Office for Europe "Health Effects of Particulate Matter: Policy Implications for Countries in Eastern Europe, Caucasus and Central Asia." Copenhagen, Denmark: WHO Regional Office for Europe, 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.

Yoon, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L. Shapley. "Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement." *Issues & Answers Report*. Washington, DC: Regional Education Laboratory Southwest, 2007.

This page has been left blank for double-sided copying.

Appendix A

ILEI evaluation intention-to-treat impact (ITT) impact estimates

This page has been left blank for double-sided copying.

As discussed in Chapter II, we estimated the impacts of the school rehabilitation activity using an intent-to-treat (ITT) model as well as treatment-on-the-treated (TOT) model. In this appendix we present the intent-to-treat estimates showing the average effect of being *assigned* to receive rehabilitation (as opposed to the effect of rehabilitation itself). The ITT model used in this appendix requires fewer assumptions for the impact estimates to be valid: specifically, unlike the TOT model the ITT model does not require an assumption that school rehabilitation is the only way assignment to the treatment group or control group could have affected the evaluation's outcomes of interest.

In practice, the ITT model and TOT models both produced a very similar pattern of impacts on school infrastructure outcomes: rehabilitation produced a dramatic improvement in overall building quality, heating systems and heating-system usage, air quality, sanitation facilities, and perceptions of school safety and the perceived quality of the learning environment among students and teachers.

A. Effects on school infrastructure

1. Physical condition of the school building

Similar to the results from the study's TOT model, the ITT analysis showed that assignment to the treatment group produced substantial improvements in the physical infrastructure of rehabilitated schools. Treatment schools experienced large and statistically significant improvements in 1) **the overall condition of the school building, 2) the condition of the walls, ceilings, and floors, and 3) the** condition of indoor stairs in the main school building (Table A.1). Assignment to the treatment group also greatly reduced the percentage of schools that had a classroom with three or more problems with walls (-69 percentage points), ceilings (-70 percentage points,), or floors (-55 percentage points), as shown in Table A.2. Assignment to the treatment group also greatly increased the percentage of teachers and students from treatment schools who reported being satisfied with the quality of the building and equipment at their schools (difference > 50 percentage points). Assignment to treatment also substantially reduced the percentage of teachers and students who reported that their school needed immediate repairs. (Table A.3).

Table A.1. Intention-to-Treat (ITT) impact of rehabilitation on school infrastructure and teaching facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Better condition of school building exterior (z-score)	0.13 s.d	-0.15 s.d.	0.28^ s.d.	0.096	0.166
Better condition of walls, ceilings, and floors in all classrooms and indoor gym (z-score)	0.72 s.d.	-0.85 s.d.	1.57** s.d.	0.000	0.089
Better condition of stairs in main school building (z-score)	0.59 s.d.	-0.71 s.d.	1.31** s.d.	0.000	0.116
School has an indoor gym (p.p.)	84.1	78.9	5.2	0.366	0.057
School has an outdoor recreation area (p.p.)	64.5	72.7	-7.8	0.271	0.070
School has a science laboratory (p.p.)	91.3	31.2	60.2**	0.000	0.060

Source: Two-year follow-up building survey administered in 175 schools.

Notes: The first three rows show data on building condition indices with a mean of zero and a standard deviation of 1 (z-scores), with means and differences calculated in standard deviation (s.d.) units.

Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.2. Intention-to-Treat (ITT) impact of rehabilitation on problematic conditions in walls, ceiling, and floors of classrooms

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of schools that have at least one classroom with:					
Three or more problems in walls	11.0	79.9	-68.8**	0.000	0.056
Three or more problems in ceilings	2.3	72.6	-70.3**	0.000	0.054
Three or more problems in floors	6.2	60.8	-54.6**	0.000	0.058

Source: Two-year follow-up building survey administered in 175 schools.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.3. Intention-to-Treat (ITT) impact of rehabilitation on student/teacher perceptions related to the quality of physical building

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers					
Satisfied or very satisfied with the quality of the building and equipment at their school	93.3	32.7	60.6**	0.000	0.041
Reported that school needs immediate repairs or improvements	15.9	84.5	-68.6**	0.000	0.043
Students					
Satisfied or very satisfied with the quality of the building and equipment at their school	81.0	28.0	53.0**	0.000	0.027
Reported that school needs immediate repairs or improvements	13.7	82.8	-69.1**	0.000	0.028

Source: Teacher and student surveys completed by 1,385 teachers and 7,863 students, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

2. Effects on heating systems

As with the TOT model which was the focus on the main report, the ITT impact estimates show that assignment to the treatment group increased the availability of central heating systems in schools, increased the availability of central heating in all classrooms and indoor gyms, and improved classroom temperatures in winter month (Table A.4). Students, teachers, and parents felt that improved heating systems enhanced the learning environment in school (Table A.5).

Table A.4. Intention-to-Treat (ITT) impact of rehabilitation on presence and perception of central heating

	Treatment (A)	Control (B)	Impact (A-B)	p-value	Standard error
Have functional central heating	95.6	62.1	-33.5**	0.000	0.055
All classrooms have functional central heating	91.4	62.1	29.3**	0.000	0.059
Indoor gym has central heating	99.2	64.3	34.9**	0.000	0.061
Average measured temperature (median classroom, degrees in Celsius)	18.9	16.0	2.9**	0.000	0.472

Source: Building survey administered in 175 schools at two-year follow up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.5. Intention-to-Treat (ITT) impact of rehabilitation on student/teacher perceptions related to the quality of physical building

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Students					
Feels classroom is too cold, on average, in February	7.5	36.7	-29.1**	0.000	0.031
Feels temperature negatively affected ability to concentrate in February	13.0	37.0	-24.0**	0.000	0.025
Teachers					
Feels classroom is too cold, on average, in February	1.2	27.7	-26.5**	0.000	0.039
Feels temperature negatively affected ability to teach in February	3.9	18.7	-14.8**	0.000	0.032
Parents					
Feels classroom is too cold, on average, in February	4.90	30.79	-25.89**	0.000	0.034

Source: Teacher, parent, and student surveys completed by 1,385 teachers, 6,609 parents, and 7,863 students, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

3. Effects on air quality outcomes

As with the TOT model, ITT impact estimates show that assignment to the treatment group dramatically improved air quality measured as exposure to $PM_{2.5}$ and PM_{10} levels. The average $PM_{2.5}$ and PM_{10} levels values in rehabilitated schools were markedly lower than control schools. These findings were not driven by the most extreme classrooms with the worst air quality in each school: we also found similarly large improvements in air quality for the median classroom (Table A.6). The program helped rehabilitated schools to meet WHO air-quality guidelines (Table A.7), and substantially reduced the percentage of students and teachers who reported that air quality in the classroom in past month was poor or unhealthy. Similarly, the program reduced the percentage of students and teachers who reported that air quality affected instruction during cold weather (Table A.8).

Table A.6. Intention-to-Treat (ITT) impact of rehabilitation on air quality outcomes

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Average $PM_{2.5}$ (max classroom)	20.1	79.4	-59.2**	0.000	13.017
Average PM_{10} (max classroom)	40.3	117.1	-76.8**	0.000	16.285
Average $PM_{2.5}$ (median classroom)	13.2	57.8	-44.6**	0.000	11.219
Average PM_{10} (median classroom)	26.7	111.8	-85.1**	0.000	22.223

Source: Building survey administered in 175 schools at two-year follow up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.7. Intention-to-Treat (ITT) impact on percentage of schools meeting WHO interim air quality targets ($PM_{2.5}$ and PM_{10})

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of schools meeting WHO interim air quality targets					
Average $PM_{2.5}$	85.6	61.2	24.4**	0.000	0.066
Average PM_{10}	84.5	62.5	22.0**	0.001	0.068

Source: Building survey administered in 175 schools at two-year follow up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.8. Intention-to-Treat (ITT) impact of rehabilitation on perceived air quality in schools in February

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Students					
Air quality in classroom in past month was poor or unhealthy	14.8	40.2	-25.4**	0.000	0.025
Air quality affected student's ability to concentrate in Feb	13.5	31.5	-17.9**	0.000	0.020
Air quality disrupted instruction in February	9.5	18.7	-9.3**	0.000	0.017
Teachers					
Air quality in classroom in past month was poor or unhealthy	1.7	22.5	-20.8**	0.000	0.030
Air quality disrupted instruction in February	2.7	14.9	-12.2**	0.000	0.026

Source: Teacher and student surveys completed by 1,385 teachers and 7,863 students, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

4. Effects on lighting and electrical systems

As with the TOT model, the ITT impact estimates show that most of the classrooms in treatment schools had functional electric lighting, whereas about half of control schools had at least one classroom without lighting. More control than treatment students reported having problems reading because of lighting (Table A.9).

Table A.9. Intention-to-Treat (ITT) impact of rehabilitation on lighting and its effect on the learning environment

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools					
At least one classroom without working lighting in school	6.4	50.7	-44.3**	0.000	0.056
Students					
Ever have difficulty reading because of lighting	6.5	36.2	-29.8**	0.000	0.026
Ever have difficulty reading blackboard because of lighting	36.0	63.7	-27.8**	0.000	0.022
Feels lighting negatively affected ability to concentrate on schoolwork in February	6.2	27.9	-21.7**	0.000	0.023
Teachers					
Feels lighting is insufficient for students	7.7	35.4	-27.7**	0.000	0.042

Source: Teacher and student surveys completed by 1,385 teachers and 7,863 students, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

5. Effects on outcomes related to sanitary facilities

The ITT model also produced very similar results to the TOT model with respect to outcomes related to sanitary facilities. Assignment to the treatment group reduced the percentage of schools without a functional toilet by 15 percentage points. The program also improved sanitary conditions and cleanliness of toilet facilities in treatment schools (Table A.10). The intervention also substantially improved student comfort using sanitary facilities (Table A.11). This change was driven by a dramatic improvement in the percentage of students reporting that they “always” felt comfortable using sanitary facilities in treatment schools. Improvements in comfort using sanitary facilities were similar among female and male students. While the differences between treatment and control students were statistically significant for female and male students, results for both genders are statistically indistinguishable from one another (Table A.12).

Table A.10. Intention-to-Treat (ITT) impact of rehabilitation on sanitary facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools without					
Flushing toilet	18.8	33.7	-14.9*	0.029	0.067
Running water for hand washing	7.6	25.7	-18.1**	0.001	0.054
Soap near toilets or latrines	23.1	42.3	-19.3**	0.004	0.066
Schools with an odor in restroom facilities	25.4	70.5	-45.1**	0.000	0.061

Source: Building survey administered in 175 schools at two-year follow up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.11. Intention-to-Treat (ITT) impact of rehabilitation on student comfort using sanitary facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of students who reported feeling comfortable using sanitary facilities:					
Always	57.6	26.1	31.5**	0.000	0.028
Sometimes	18.5	17.7	0.8	0.517	0.012
Rarely	10.9	18.2	-7.3**	0.000	0.011
Never	13.0	38.0	-25.0**	0.000	0.025

Source: Student surveys completed by 7,863 students, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.12. Intention-to-Treat (ITT) impact of rehabilitation on student comfort using sanitary facilities by gender

	Female students		Male students		Difference in impacts by gender (p-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Students who reported they were					
Always comfortable using the sanitary facilities	32.5**	0.000	30.6	0.000	0.497
Sometimes comfortable using the sanitary facilities	0.7	0.694	0.9	0.537	0.940
Rarely comfortable using the sanitary facilities	-8.4**	0.000	-6.3	0.000	0.211
Never comfortable using the sanitary facilities	-24.8**	0.000	-25.1	0.000	0.895

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

B. Effects on instructional time, facility use, and school safety

1. Instructional time

As with the TOT model, impact estimates from the ITT model did not reveal differences between treatment and control schools in the absenteeism patterns reported by teachers. We also estimated the percentage of enrolled students absent on an average day and found no differences between treatment and control groups (Table A.13). The program also did not affect the amount of time teachers report spending on classroom instruction (Table A.14).

Table A.13. Intention-to-Treat (ITT) impact of rehabilitation on teachers reported students' absences

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers reported students' absences on an average day in the previous month:					
Perfect attendance	12.5	10.2	2.3	0.373	0.026
Between one and four students absent	57.8	58.1	-0.4	0.929	0.041
Five or more students absent	29.7	31.7	-2.0	0.636	0.042
Percentage of enrolled students absent on average day	16.7	17.2	-0.5	0.650	0.011

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table A.14. Intention-to-Treat (ITT) impact of rehabilitation on class time spent on instruction per day in the month

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Class time spent on instruction per day in the month before the two-year follow-up surveys					
less than an hour	4.8	3.3	1.4	0.210	0.011
one to two hours	36.3	38.9	-2.6	0.536	0.042
three to four hours	41.4	35.7	5.7	0.135	0.038
five or more hours	17.6	22.1	-4.5	0.158	0.032

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the percent 10/5/1 percent levels.

2. Use of science labs

As with the TOT model, the ITT analysis also showed that assignment to the treatment group increased the percentage of treatment schools with a science laboratory and the availability of science equipment such as microscope, heating devices, chemicals, lab coats, protective eyewear, and beakers. The program also produced significant improvements in students' exposure to science laboratories, including receiving more science demonstrations and participating in experiments (Table A.15).

Table A.15. Intention-to-Treat (ITT) impact of rehabilitation on students' exposure to science laboratories

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers					
Always/sometimes shown demonstrations (if teaching science)	**76.6**	**19.4**	**57.21****	**0.000**	**0.041**
Always/sometimes conduct own experiments (if teaching science)	**73.4**	**13.6**	**59.89****	**0.000**	**0.040**
Students					
School has a science laboratory	94.9	39.1	55.8**	0.000	0.035
Students who reported teacher always/sometimes shown demonstrations	55.0	23.8	31.3**	0.000	0.029
Students always/sometimes conduct own experiments (if teachers did demos)	52.2	26.3	25.9**	0.000	0.024
Availability of science equipment (if teachers did demos)					
Equipment not available in science class	0.4	8.1	-7.7**	0.000	0.013
Microscope	75.9	55.8	20.1**	0.000	0.030
Heating devices	47.5	18.6	28.9**	0.000	0.024
Chemicals or other materials for experiments	63.6	34.9	28.8**	0.000	0.032
Lab coats	44.5	8.6	35.9**	0.000	0.027
Protective eyewear	49.2	10.9	38.3**	0.000	0.026
Beakers	63.8	50.5	13.8**	0.000	0.025
Other science equipment	1.3	1.1	0.2	0.428	0.003
Availability of services or electronic equipment (if teachers did demos)					
Computer	48.2	48.7	-0.5	0.877	0.030
Internet access	40.1	24.8	15.4**	0.000	0.030
Projection screen	36.9	28.5	8.4**	0.003	0.028
Television	7.0	7.3	-0.3	0.902	0.022
Other electronic equipment	1.6	1.2	0.4	0.315	0.004

Source: Teacher surveys completed by 532 science teachers and 5.444 students who received science demonstrations, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

3. Use of recreational facilities

As with the TOT results, the ITT analysis showed that most schools had at least some type of indoor gym prior to the intervention; however, the program modestly increased the availability of indoor gyms, and increased the usage of indoor recreational activities (Table A.16).

Table A.16. Intention-to-Treat (ITT) impact of rehabilitation on use of recreational school facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Availability of indoor gym	86.1	79.3	6.8^	0.084	0.039
Student reported using at least once in an average week					
Indoor gym (if available)	92.3	86.5	5.9*	0.013	0.023
Outdoor recreation area (if available)	68.8	65.9	2.8	0.501	0.042

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

4. School safety

Similar to the results from the TOT analysis, ITT impact estimates show that the program substantially improved the percentage of respondents (students, their parents, teachers, and school directors) who felt safe in the school. More students, parents, teachers, and school directors from treatment than control schools reported that school facilities were safe (Table A.17).

Table A.17. Intention-to-Treat (ITT) impact of rehabilitation on perceived safety

	Treatment (A)	Control (B)	Impact (A-B)	p-value	Standard error
Students					
Agrees that the school is safe and healthy	84.8	43.9	40.9**	0.000	0.028
Feels very safe in the classroom	89.0	51.4	37.6**	0.000	0.028
Feels very safe using stairwells	85.6	43.0	42.6**	0.000	0.030
Parents					
Agrees that the school is safe and healthy	94.5	60.7	33.9**	0.000	0.030
Feels that students are very safe in the classroom	90.9	47.3	43.7**	0.000	0.032
Feels that stairwells are very safe	89.4	47.0	42.3**	0.000	0.035
Teachers					
Agrees that the school is safe	97.9	64.5	33.4**	0.000	0.041
Agrees that the school is healthy	97.9	70.7	27.2**	0.000	0.041
Feels very safe in the classroom	97.2	65.9	31.3**	0.000	0.039
Feels that students are very safe in the classroom	96.9	65.8	31.1**	0.000	0.040
Feels very safe using stairwells	96.3	55.6	40.7**	0.000	0.041
Feels that students are very safe using stairwells	97.9	63.4	34.5**	0.000	0.041
School directors					
Agrees that the school is safe	94.4	56.1	38.3**	0.000	0.065

Source: Teacher surveys completed by 1,376 teachers, interviewed at two-year follow-up.

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

C. Effects on learning outcomes

As with the TOT impact analysis, ITT impact estimates showed that student test scores in language, math, and science were very similar in treatment and control schools. We did not find evidence that the program increased test scores in upper-secondary grades, two years after rehabilitation was completed (Table A.18).

Table A.18. Intention-to-Treat (ITT) impact of rehabilitation on student test scores across grades 10, 11, and 12

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Language	0.06	0.04	0.01	0.812	0.063
Math	0.05	0.03	0.02	0.766	0.071
Science	0.09	0.16	-0.07	0.297	0.068

Source: Administrative data from the Education Management Information System (EMIS).

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Similar to the results from the TOT model, the ITT analysis showed that assignment to the treatment group appears to have had a positive effect on learning outcomes in language and math before the pandemic, but these effects were absent for schools assessed after the pandemic (Table A.19). However, the sample sizes in pre-pandemic and post-pandemic groups of schools are not large enough to detect if the impacts in the two periods are statistically significant: in addition, the *difference* between pre- and post-pandemic impacts was only marginally statistically significant (at the 10 percent level) for language test scores, and not significant for math test scores. In science, the pre-pandemic effect of rehabilitation was close to zero, and the impact on science scores became more-markedly negative after the pandemic. The difference in impacts on science score across the two periods was not statistically significant.

Table A.19. Intention-to-Treat (ITT) impact of rehabilitation on student test scores across grades 10, 11, and 12

	Before the pandemic		After the pandemic		
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	Difference in impacts (p-value)
Language	0.12	0.202	-0.09	0.290	0.095
Math	0.09	0.388	-0.04	0.725	0.417
Science	-0.03	0.816	-0.12	0.271	0.556

Source: Administrative data from the Education Management Information System (EMIS).

Notes: Columns A and B present ordinary least-squares (OLS) regression-adjusted group means. OLS estimates the average effect of being assigned to receive school rehabilitation. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Appendix B

ILEI evaluation impact estimates one year after rehabilitation

This page has been left blank for double-sided copying.

In this appendix we present the treatment-on-the-treated (TOT) estimates for the ILEI evaluation one year after rehabilitation was completed, for each of the outcomes presented in the report (the primary analyses in the main report focused on outcomes two years after rehabilitation, rather than the one-year follow-up). Broadly speaking, the impact estimates after one year look very similar to the impact estimates observed after two years. By the one-year follow up, rehabilitation produced a dramatic improvement in overall building quality, heating systems and heating-system usage, air quality, sanitation facilities, and perceptions of school safety and the perceived quality of the learning environment among students and teachers.

A. Effects on school infrastructure

1. Physical condition of the school building

By the one-year follow up, the program produced substantial improvements in the physical infrastructure of rehabilitated schools (Table B.1). Rehabilitation greatly reduced the percentage of schools that had a classroom with three or more problems with walls (-81 percentage points), ceilings (-74 percentage points), or floors (-78 percentage points), as shown in Table B.2. Rehabilitation also greatly improved the percentage of teachers (by 62 percentage points) and students (by 49 percentage points) who reported that they are very satisfied with the quality of the building and equipment at their schools. Rehabilitation also greatly reduced the percentage of teachers and students who reported that their school needed immediate repairs (Table B.3).

Table B.1. One-year impact of rehabilitation on school infrastructure and teaching facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Better condition of school building exterior (z-score)	0.65 s.d	-0.63 s.d.	1.28** s.d.	0.000	0.129
Better condition of walls, ceilings, and floors in all classrooms and indoor gym (z-score)	0.89 s.d.	-0.90 s.d.	1.78** s.d.	0.000	0.079
Better condition of stairs in main school building (z-score)	0.96 s.d.	-0.73 s.d.	1.70** s.d.	0.000	0.089
School has an indoor gym (p.p.)	85.4	77.5	7.9	0.208	0.063
School has an outdoor recreation area (p.p.)	61.0	74.1	-13.1^	0.063	0.071
School has a science laboratory (p.p.)	97.9	34.6	63.3**	0.000	0.062

Source: One-year follow-up building survey administered in 165 schools. For the one-year follow-up 11 schools were closed or refused to participate in the building survey.

Notes: The first three rows show data on building condition indices with a mean of zero and a standard deviation of 1 (z-scores), with means and differences calculated in standard deviation (s.d.) units.

Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table B.2. One-year impact of rehabilitation on problematic conditions in walls, ceiling, and floors of classrooms

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of schools that have at least one classroom with:					
Three or more problems in walls	4.6	85.2	-80.5**	0.000	0.052
Three or more problems in ceilings	5.7	79.3	-73.6**	0.000	0.055
Three or more problems in floors	0.0	78.8	-78.8**	0.000	0.050

Source: One-year follow-up building survey administered in 165 schools. For the one-year follow-up 11 schools were closed or refused to participate in the building survey.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table B.3. One-year impact of rehabilitation on student/teacher perceptions related to the quality of physical building

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers					
Satisfied or very satisfied with the quality of the building and equipment at their school	90.6	28.6	61.9**	0.000	0.116
Reported that school needs immediate repairs or improvements	12.4	92.1	-79.7**	0.000	0.077
Students					
Satisfied or very satisfied with the quality of the building and equipment at their school	80.5	31.8	48.7**	0.000	0.093
Reported that school needs immediate repairs or improvements	14.1	84.7	-70.6**	0.000	0.076

Source: Teacher and student surveys completed by 1,397 teachers and 9,080 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

2. Effects on heating systems

By the one-year follow up, rehabilitation increased the availability of central heating systems in schools (about 44 percentage points). It also increased the availability of central heating in all classrooms (48 percentage points) and indoor gyms (42 percentage points). Rehabilitation also improved classroom temperatures in winter months (Table B.4). As shown in Table B.5., students, teachers, and parents felt that improved heating systems improved the learning environment in school.

Table B.4. One-year impact of rehabilitation on presence and perception of central heating

	Treatment (A)	Control (B)	Impact (A-B)	p-value	Standard error
Have functional central heating	100.0	56.4	43.6**	0.000	0.055
All classrooms have functional central heating	97.7	49.6	48.1**	0.000	0.065
Indoor gym has central heating	92.8	51.0	41.8**	0.000	0.076
Average measured temperature (median classroom, degrees in Celsius)	20.0	16.9	3.1**	0.000	0.435

Source: Building survey administered in 165 schools at one-year follow up. For the one-year follow-up 11 schools were closed or refused to participate in the building survey.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table B.5. One-year impact of rehabilitation on student/teacher perceptions related to the quality of physical building

	Treatment (A)	Control (B)	Impact (A-B)	p-value	Standard error
Students					
Feels classroom is too cold, on average, in February	0.6	44.3	-43.7**	0.000	0.110
Feels temperature negatively affected ability to concentrate in February	11.3	40.8	-29.5**	0.003	0.098
Teachers					
Feels classroom is too cold, on average, in February	0.0	25.0	-25.0*	0.025	0.117
Feels temperature negatively affected ability to teach in February	4.5	15.6	-11.1	0.160	0.079
Parents					
Feels classroom is too cold, on average, in February	0.7	30.9	-30.1**	0.001	0.088

Source: Teacher and student surveys completed by 1,397 teachers and 8,728 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

3. Effects on air quality outcomes

After one year, rehabilitation dramatically improved air quality measured as exposure to $PM_{2.5}$ and PM_{10} levels. The average $PM_{2.5}$ and PM_{10} levels values in rehabilitated schools were markedly lower than control schools. These findings were not driven by the most extreme classrooms with the worst air quality in each school: we also found similarly large improvements in air quality for the median classroom (Table B.6). By the one-year follow up, rehabilitation also helped schools meet the WHO's interim guidelines for $PM_{2.5}$ and PM_{10} exposure (Table B.7), and substantially reduced the percentage of students and teachers who reported that air quality in the classroom in past month was poor or unhealthy (Table B.8).

Table B.6. One-year impact of rehabilitation on air quality outcomes

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Average $PM_{2.5}$ (max classroom)	16.2	85.1	-68.9**	0.000	13.485
Average PM_{10} (max classroom)	35.6	131.8	-96.1**	0.000	17.143
Average $PM_{2.5}$ (median classroom)	12.8	54.9	-42.1**	0.000	10.639
Average PM_{10} (median classroom)	28.1	100.7	-72.7**	0.000	18.494

Source: Building survey administered in 165 schools at one-year follow up. For the one-year follow-up 11 schools were closed or refused to participate in the building survey.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table B.7. One-year impact of schools meeting WHO interim air quality targets ($PM_{2.5}$ and PM_{10})

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of schools meeting WHO interim air quality targets					
Average $PM_{2.5}$	89.8	61.9	27.9**	0.000	0.066
Average PM_{10}	87.6	60.7	26.9**	0.000	0.067

Source: Building survey administered in 165 schools at one-year follow up. For the one-year follow-up 11 schools were closed or refused to participate in the building survey.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table B.8. One-year impact of rehabilitation on perceived air quality in schools in February

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Students					
Air quality in classroom in past month was poor or unhealthy	16.4	42.6	-26.2**	0.009	0.101
Air quality affected student's ability to concentrate in Feb	10.0	39.9	-30.0**	0.001	0.092
Air quality disrupted instruction in February	5.2	23.9	-18.7**	0.009	0.071
Teachers					
Air quality in classroom in past month was poor or unhealthy	0.0	33.2	-33.2**	0.001	0.115
Air quality disrupted instruction in February	3.2	17.1	-13.9^	0.094	0.083

Source: Teacher and student surveys completed by 1,397 teachers and 8,468 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

4. Effects on lighting and electrical systems

After one year, rehabilitation produced a large improvement in the availability of functional electric lighting. More control than treatment students also reported having problems reading in classroom due to insufficient lighting (Table B.9).

Table B.9. One-year impact of rehabilitation on lighting and its effect on the learning environment

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools					
At least one classroom without working lighting in school	12.4	63.3	-50.82**	0.000	0.071
Students					
Ever have difficulty reading because of lighting	7.2	31.4	-24.2**	0.004	0.084
Ever have difficulty reading blackboard because of lighting	34.7	66.3	-31.7**	0.000	0.073
Feels lighting negatively affected ability to concentrate on schoolwork in February	5.5	26.3	-20.9**	0.006	0.076
Teachers					
Feels lighting is insufficient for students	10.2	33.5	-23.3*	0.039	0.113

Source: Teacher and student surveys completed by 1,390 teachers and 9,247 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

5. Effects on sanitation outcomes

After one year, the program reduced the percentage of schools without a functional flushing toilet by 47 percentage points. The program also improved sanitary conditions and cleanliness of toilet facilities in treatment schools (Table B.10). These changes in rehabilitated schools substantially improved the percentage of students who reported "always" feeling comfortable using sanitary facilities (Table B.11). Improvements in comfort using sanitary facilities were similar among female and male students. While the differences between treatment and control students were statistically significant for both male and female students, but results for both genders were statistically indistinguishable from one another one year after rehabilitation (Table B.12).

Table B.10. One-year impact of rehabilitation on sanitary facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools without					
Flushing toilet	0.9	48.0	-47.1**	0.000	0.061
Running water for hand washing	8.1	36.4	-28.24**	0.000	0.067
Soap near toilets or latrines	27.8	52.7	-24.96**	0.002	0.079
Schools had an odor in restroom facilities	25.4	79.2	-53.72**	0.000	0.070

Source: Building survey administered in 165 schools at one-year follow up. For the one-year follow-up 11 schools were closed or refused to participate in the building survey.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

Table B.11. One-year impact of rehabilitation on student comfort using sanitary facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of students who reported feeling comfortable using sanitary facilities:					
Always	62.8	13.5	49.3**	0.000	0.107
Sometimes	15.3	19.9	-4.6	0.478	0.065
Rarely	8.3	19.6	-11.3*	0.026	0.051
Never	13.5	47.0	-33.5**	0.009	0.127

Source: Student surveys completed by 8,876 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

Table B.12. One-year impact of rehabilitation on student comfort using sanitary facilities by gender

	Female students		Male students		Difference in impacts by gender (p-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Students who reported they were					
Always comfortable using the sanitary facilities	45.6**	0.000	52.5**	0.006	0.717
Sometimes comfortable using the sanitary facilities	3.8	0.039	-11.8	0.321	0.197
Rarely comfortable using the sanitary facilities	-5.8**	0.001	-16.0^	0.087	0.287
Never comfortable using the sanitary facilities	-43.6**	0.000	-24.8	0.288	0.422

Source: Student surveys completed by 8,876 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the percent 10/5/1 levels.

B. Effects on instructional time, facility use, and school safety

1. Instructional time

One year after rehabilitation, there were no differences between treatment and control schools in the absenteeism patterns reported by teachers. We also estimated the percentage of enrolled students absent on an average day and found no differences between treatment and control groups (Table B.13). The program also did not affect the amount of time teachers report spending on classroom instruction after one year (Table B.14).

Table B.13. One-year impact of rehabilitation on teachers reported students' absences

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers reported students' absences on an average day in the previous month:					
Perfect attendance	20.5	11.6	8.9	0.323	0.090
Between one and four students absent	65.4	76.2	-10.8	0.305	0.105
Five or more students absent	14.2	11.9	2.3	0.790	0.087
Percentage of enrolled students absent on average day	10.3	12.8	-2.4	0.356	0.026

Source: Teacher surveys completed by 1,385 teachers, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

Table B.14. One-year impact of rehabilitation on class time spent on instruction per day in the month

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Class time spent on instruction per day in the month before the two-year follow-up surveys					
less than an hour	6.9	6.5	0.5	0.931	0.053
one to two hours	36.3	43.6	-7.2	0.619	0.146
three to four hours	40.1	28.1	12.0	0.360	0.131
five or more hours	16.6	21.8	-5.2	0.695	0.133

Source: Teacher surveys completed by 1,396 teachers, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 levels.

2. Use of science labs

The intervention increased the percentage of treatment schools with a science laboratory and the availability of science equipment such as microscope, heating devices, chemicals, lab coats, protective eyewear, and beakers. By the one-year follow-up, the program also produced significant improvements in students' exposure to science laboratories, including receiving more science demonstrations and participating in experiments (Table B.15).

Table B.15. One-year impact of rehabilitation on students' exposure to science laboratories

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Teachers					
Always/sometimes shown demonstrations (if teaching science)	74.4	43.5	30.9*	0.021	0.134
Always/sometimes conduct own experiments (if teaching science)	71.1	19.7	51.4**	0.000	0.140
Students					
School has a science laboratory	97.9	41.3	56.7**	0.000	0.124
Students who reported teacher always/sometimes shown demonstrations	48.4	37.2	11.2	0.464	0.153
Students always/sometimes conduct own experiments (if teachers did demos)	42.6	42.2	0.4	0.981	0.166
Availability of science equipment (if teachers did demos)					
Equipment not available in science class	0.0	15.5	-15.5*	0.017	0.075
Microscope	73.1	48.5	24.5*	0.035	0.117
Heating devices	42.5	18.2	24.3*	0.011	0.096
Chemicals or other materials for experiments	65.3	33.1	32.1**	0.008	0.120

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Lab coats	41.7	5.9	35.8**	0.000	0.090
Protective eyewear	43.8	12.2	31.6**	0.000	0.087
Beakers	63.2	52.4	10.8	0.192	0.083
Other science equipment	1.3	1.0	0.2	0.809	0.010
Availability of services or electronic equipment (if teachers did demos)					
Computer	38.9	52.9	-14.0	0.279	0.129
Internet access	24.0	33.9	-9.9	0.405	0.119
Projection screen	25.3	39.0	-13.7	0.238	0.116
Television	6.3	4.2	2.0	0.868	0.122
Other electronic equipment	3.3	0.9	2.4	0.418	0.030

Source: Teacher surveys completed by 575 science teachers and 6,448 students who received science demonstrations, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

3. Use of recreational facilities

By the one-year follow-up, the program increased the availability of indoor gyms but the difference (14 percentage points) was not statistically significant. We also found no differences in the percentage of students reporting that they used indoor gyms and outdoor recreational spaces (Table B.16).

Table B.16. One-year impact of rehabilitation on use of recreational school facilities

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Availability of indoor gym	91.6	77.5	14.1	0.310	0.139
Student reported using at least once in an average week					
indoor gym (if available)	96.3	86.3	10.0	0.256	0.088
outdoor recreation area (if available)	59.2	80.3	-21.0	0.233	0.177

Source: Student surveys completed by 7,978 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

4. School safety

One year after rehabilitation, students, their parents, teachers, and school directors all reported that rehabilitation improved safety in schools. Rehabilitation substantially improved the percentage of respondents who felt safe in the school (Table B.17). Rehabilitation also improved the percent of students and teachers reporting feeling very safe in the classroom and while using the school's stairwells.

Table B.17. One-year impact of rehabilitation on perceived safety

	Treatment (A)	Control (B)	Impact (A-B)	p-value	Standard error
Students					
Agrees that the school is safe and healthy	84.4	46.4	38.0**	0.000	0.095
Feels very safe in the classroom	90.7	48.2	42.5**	0.000	0.109
Feels very safe using stairwells	83.9	52.2	31.6**	0.002	0.102
Parents					
Agrees that the school is safe and healthy	95.9	60.4	35.5**	0.000	0.087
Feels that students are very safe in the classroom	88.5	46.5	42.0**	0.000	0.104
Feels that stairwells are very safe	88.0	46.6	41.4**	0.000	0.095
Teachers					
Agrees that the school is safe	92.9	66.5	26.4*	0.033	0.124
Agrees that the school is healthy	94.3	70.0	24.3*	0.033	0.114
Feels very safe in the classroom	87.3	78.6	8.8	0.553	0.148
Feels that students are very safe in the classroom	88.8	75.8	13.0	0.368	0.145
Feels very safe using stairwells	91.2	56.4	34.8*	0.011	0.138
Feels that students are very safe using stairwells	87.2	76.1	11.1	0.428	0.140
School directors					
Agrees that the school is safe	96.6	51.4	45.2**	0.000	0.064

Source: Student, parent, teachers and school director surveys completed by 9,162 students, 7,579 parents, 1,369 teachers, and 167 school directors, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

C. Effects on learning outcomes

One year after rehabilitation, TOT impact estimates showed that student test scores in language and math were very similar in treatment and control schools. However, the activity increased test scores in science one year after rehabilitation was completed (Table B.18). This finding of positive learning effects after one year is consistent with results in the main report showing that the impacts of school rehabilitation on learning outcomes were more positive prior to the COVID-19 pandemic. Compared to the two-year follow-up impact estimates that are the focus of the main report, the study collected one-year outcome data prior to the pandemic for a substantially larger proportion of the sample (171 out of 176 schools for language and 158 out of 176 schools for math and science).

Table B.18. One-year impact of rehabilitation on student test scores across grades 10, 11, and 12

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Language	0.02	-0.03	0.05	0.559	0.080
Math	0.00	-0.07	0.07	0.417	0.087
Science	0.05	-0.13	0.18*	0.037	0.086

Source: Education Management Information System (EMIS) administrative data from 171 schools for language and 158 schools for math and science scores, and student samples sizes ranging from 7,739 to 8,662.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program.. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

This page has been left blank for double-sided copying.

Appendix C

Summary of qualitative findings for the ILEI evaluation

This page has been left blank for double-sided copying.

In this appendix, we present a summary of the findings of the qualitative analyses we conducted for the ILEI study. Table C.1. presents (1) key qualitative findings mentioned in the report, (2) the extent to which these findings reflect statements made by students, teachers, and directors (or triangulation of findings); and (3) illustrative quotes that substantiate the findings.

This analysis draws upon stakeholder interviews at 15 rehabilitated schools and 5 control-group schools conducted in Spring 2022. At each school, researchers interviewed the school director and four teachers, and held two student focus groups (one with 10th graders and one with 12th graders). For more information on the qualitative data sources and analysis methods, see Chapter II of this report (Section II.4).

Table C.1. Summary of qualitative findings

Key findings from qualitative data	Triangulation of findings by stakeholder			*Examples of qualitative evidence: Illustrative quotes*
	Students' perceptions	Teachers' perceptions	Directors' perceptions	
Physical condition of the school building				
Addressing highly visible infrastructure problems was important to students and teachers. (Page 30)	X	X	X	*"This firewood was sometimes wet, sometimes it was windy, the stove pipes didn't fit properly in the windows. The wind was blowing inside, sometimes we were in smoke, we were smoked…, it was horrible, it was hell. Now we now have 21st century heating."* *– Director rehabilitated school in Samtskhe-Javakheti*
Effects on heating systems				
Improved heating systems improved the learning environment in school. (Page 26)	X	X	X	*"We feel more comfortable obviously and we do not have to worry about the cold or the wind blowing through the broken glass now. The physical environment has improved quite a bit, not even improved, it has radically changed, and this will obviously be reflected on the learning process as well."* *– Director, rehabilitated school in Guria*
Effects on air quality outcomes				
Air quality improvements helped to improve students' time on task. (Page 31)	X	X	X	*"[Before] when we were using the wood stove, on windy days the smoke was coming inside and we were in tears all the time and we were walking around the corridors looking for a place where we could hold lessons. This has changed, and now we do not spend any time on ventilation and stove fuel collection and so on. I used to have to go outside to breathe air before, it does not happen anymore. The lesson process is more organized"* *– STEM Teacher, rehabilitated school in Guria*
Effects on lighting and electrical systems				
Teachers and students benefited from improved electrical wiring and access to outlets. (Page 33)	X	X	X	*"Previously we did not have electricity at all. We only had electricity in rooms which were used for administrative purposes, because you need internet, you need a computer, and it was transmitted through wires and we could not have electricity in classrooms. And the children could not enjoy it in any way. The internet has improved now because we have Wi- fi all over the school."* *– Director of rehabilitated school in Shida Qartli*

Key findings from qualitative data	Triangulation of findings by stakeholder			*Examples of qualitative evidence: Illustrative quotes*
	Students' perceptions	Teachers' perceptions	Directors' perceptions	
Effects on sanitation outcomes				
Improvements most appreciated by students and teachers included functional toilet facilities that were separated by gender and located inside the school. This helped teachers start lessons on time and teach with fewer interruptions. (Page 34)	X	X	X	*"[Before the rehabilitation] we had one toilet outside, a wooden toilet. Even if I had brought chlorine and some disinfectants, it would just be wasted because of the poor conditions. In addition, the whole school was relying on that one toilet, so class breaks did not give enough time for everyone to use it. Students used to go out to the toilet during lessons; this was hindering the learning process a lot. Now we have six toilets indoors, everything is well furnished, and the children think their environment is fabulous."* *– Director of rehabilitated school in Samegrelo-Zemo Svaneti*
Students in rehabilitated schools asked for further improvements in sanitary-facility maintenance and access to drinking water.	X	X	X	*"Drinking water here is the equivalent of death."* *– Student in rehabilitated school in Guria*
Instructional time				
COVID-19 may have limited the potential positive effect of infrastructure improvements on student absenteeism.	X	X	X	*"In our case, the infrastructure does not have a large impact on attendance, because the situation in the country and worldwide is larger in scale than our attendance."* *– Grade 12 students in rehabilitated school in Imereti*
The change from wood-fueled stoves to central heating has helped increase the quality of time spent on instruction. (Page 38)	X	X		*"There was a period when we used to heat with firewood and each time it was relit, we spent 15 minutes on it. Oh, how many times I used to have to light it and bring firewood from home. The classroom time use has improved a lot in this regard. The more time that you do not spend on something else, the more you can spend on learning tasks, and naturally the result is better."* *- STEM Teacher, rehabilitated school in Samegrelo-Zemo Svaneti*
Use of science labs				
Several barriers have made it difficult to access science laboratories.	X	X	X	*"We have been having a bit of a hard time lately due to the pandemic, we could not bring in children into the science lab […] we had to temporarily close these labs but we tried to bring out materials and use them in classes."* *–Director rehabilitated school in Imereti*

Key findings from qualitative data	Triangulation of findings by stakeholder			Examples of qualitative evidence: Illustrative quotes
	Students' perceptions	Teachers' perceptions	Directors' perceptions	
Capacity issues limited teachers' ability to use science labs as intensively as they had hoped. (Page 42)	X	X	X	*"I need a laboratory, I need to use the laboratory because I teach physics and math, but unfortunately, there is only one laboratory for all of the science subjects, and it is not enough for all the classes […] I try to overcome this […] But I was told I wasn't allowed to take any equipment out of the laboratory and thus everything is in there gathering dust, untouched. So, I can't take the equipment out to my rooms and I can't use the laboratory either, it would be nice if we had a solution to this problem".* *- STEM teacher from rehabilitated school in Samtskhe-Javakheti*
Access to science labs was also affected by heating issues during winter months.	X	X	X	*"There was enough science equipment for everyone, so all of us could be involved, however we had a heating issue in the lab. There are radiators, but they were never on, when we were there. There is an issue with radiators in that room too, there are only two of them. There were electric heaters for experiment purposes originally, and we have used them for warming up ourselves."* *– Grade 12 students in rehabilitated school in Mtskheta-Mtianeti,*
Use of recreational facilities				
Students feel much safer using the gym than prior to rehabilitation, and some improved gym infrastructure has increased student interest in sports.	X	X	X	*"During the rehabilitation period, the gym was also rehabilitated, there are dressing rooms, there is a doctor's room, it is a big thing for us because the children are safe."* *.- Director of rehabilitated school in Mtskheta-Mtianeti* *"My use of the gym has increased by a lot [since the rehabilitation], I am in much more athletic shape now."* *– Grade 10 students from rehabilitated school in Guria.*
School safety				
Rehabilitation has played a large role in helping to make the school feel safer. (Page 44)	X	X	X	*"I am in a safe environment, and we are not stressed about children being in danger. This is the first time when children are safer and when I'm in peace as a teacher, I contribute more to the lessons."* *– STEM teacher in rehabilitated school in Imereti*

Key findings from qualitative data	Triangulation of findings by stakeholder: Students' perceptions	Teachers' perceptions	Directors' perceptions	*Examples of qualitative evidence: Illustrative quotes*
Effects on learning outcomes				
School closures during the pandemic were especially disappointing to students and teachers in rehabilitated schools. (Page 47)	X	X	X	*"I go back to the fact that at first when we came everyone was happy, everyone was interested in even going to the lab, entering computer rooms, working directly with the computer, and then the pandemic prevented us from attending school. Not because it was an uncomfortable environment, it was because of the pandemic."* *– STEM teacher in rehabilitated school in Imereti*
Difficulties with science lab capacity and access may help to explain why the program did not improve science learning outcomes. (Page 48)	X	X	X	*"The school infrastructure helps, but the pandemic really hindered us. Two years ago, we were actively using these labs, that were put in place for us by MCA-M. But now we use them very rarely, because it is forbidden to work in a classroom without social distancing…In general, all these isolations have brought us very bad results …there really was a lot more progress and more motivation before the pandemic than in last few years."* *–STEM Teacher, rehabilitated school in Shida Qartli*
Impacts on enrollment and school administration				
Directors have been able to manage routine operating costs, but funding major repairs is a serious challenge as the infrastructure ages.		X	X	*"Maintenance is a very difficult and expensive process, as you know, we had a very harsh winter in January, especially this year, and the roof pipes of the gym were damaged, causing water to seep into the gym on all three floors and we could not solve this problem. Unfortunately, we appealed to the municipality, as we could not solve it with our own forces. There is also something in the boiler that is already damaged which I think we need to repair this summer."* *- Director from rehabilitated school in Shida Qartli*
Five years after rehabilitation, respondents also reported new challenges related to keeping heating systems and sanitation facilities in proper working order.	X	X	x	*"Maintenance is a very difficult and expensive process and we had a very harsh winter this year... the boiler is already damaged and will need repair this summer. This is difficult to fix both financially and technically. First of all, the boiler is quite expensive. Second of all, we do not have specialists in the area who could assess the damage. Separately, we often have issues with toilets – the flush breaks down very often and needs to be replaced, at least once every 2 months."* *– Director from rehabilitated school in Shida Qartli*

Key findings from qualitative data	Triangulation of findings by stakeholder			*Examples of qualitative evidence: Illustrative quotes*
	Students' perceptions	Teachers' perceptions	Directors' perceptions	
Enrollment increased noticeably at rehabilitated schools and caused some challenges related to classroom space.		X	X	*"We are asking to enlarge the building, because the number of students enrolling is increasing […] after rehabilitation we had to add more classes because parents found out that the school conditions for learning improved and started enrolling their children in our school, so we need more space. Currently, with the higher enrollment, it is hard to organize the classes especially under the pandemic regulations for social distancing. We had to move into two-shift schooling process and the parents did not like it."* *– Director from rehabilitated school in Mtskheta-Mtianeti Region*

Appendix D

ILEI evaluation impact estimates from follow-up visits in 2022

This page has been left blank for double-sided copying.

As mentioned in Chapter III, for the ILEI evaluation a final set of site visits in all study schools was conducted to complete infrastructure assessments in spring 2022. This round of data collection represented a longer-term set of follow-up visits for slightly more than half of the study schools (either three, four, or five years after rehabilitation was completed). This appendix presents the full results of these infrastructure assessments, and shows that the pattern of strong infrastructure improvements observed two years after rehabilitation remained in place in over this extended follow-up period (through spring 2022).

A. Effects on school infrastructure

1. Physical condition of the school building

In 2022, rehabilitation continued to have positive impacts on the physical infrastructure of rehabilitated schools. Treatment schools experienced large and statistically significant improvements in 1) the overall condition of the school building, 2) the condition of the walls, ceilings, and floors, and 3) the condition of indoor stairs in the main school building (Table D.1). As shown in Table D.2, across all three indices measuring school infrastructure condition, the impact estimates were larger (in magnitude) in schools that were in their third-, fourth-, or fifth-year follow-up, compared to the schools in their second follow-up year. However, none of the differences between these subgroups were statistically significant, the sample sizes in the study are only large enough to detect a very large shift in learning outcomes for subgroups of schools.

Table D.1. Impact of rehabilitation on infrastructure and teaching facilities in 2022

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Better condition of school building exterior (*z*-score)	0.6	-0.6	1.2**	0.000	0.126
Better condition of walls, ceilings, and floors in all classrooms and indoor gym (*z*-score)	0.8	-0.8	1.6**	0.000	0.107
Better condition of stairs in main school building (*z*-score)	0.8	-0.8	1.6**	0.000	0.101
School has an indoor gym (p.p.)	83.6	74.4	9.2	0.168	0.067
School has an outdoor recreation area (p.p.)	72.3	81.1	-8.8	0.176	0.065

Source: Building survey administered in 176 schools in 2022.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table D.2. Impact of rehabilitation on infrastructure and teaching facilities in 2022, by timing of the follow-up

	2022 represents the 2-year follow-up		2022 represents a longer-term follow-up		Difference in impacts by subgroup (p-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Better condition of school building exterior (*z*-score)	1.0***	0.000	1.4***	0.000	0.133
Better condition of walls, ceilings, and floors in all classrooms and indoor gym (*z*-score)	1.5***	0.000	1.6***	0.000	0.653
Better condition of stairs in main school building (*z*-score)	1.4***	0.000	1.7***	0.000	0.175
School has an indoor gym (p.p.)	-9.0	0.377	17.7**	0.039	0.047^
School has an outdoor recreation area (p.p.)	-21.2^	0.055	-2.2	0.784	0.166

Source: Building survey administered in 176 schools in 2022.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

As of spring 2022, the program also maintained its impacts in nearly eliminating problems with walls, ceilings, or floors. The percentage of schools experiencing three or more problems in walls, ceilings, and floors was less than 5 percent, compared to 60–73 percent of control schools who reported experiencing these problems (Table D.3). Reported problems with classrooms walls, ceilings, were similar among schools where 2022 represents the second-year follow-up, compared with schools that were in their third-, fourth-, or fifth-year follow-up. However, the program's impact on the percentage of schools reporting three or more problems in floors was larger for schools that were in their third, fourth, or fifth follow-up, compared to schools where 2022 represents the second follow up. The difference between this subgroup of schools was statistically significant (Table D.4).

Table D.3. Impact of rehabilitation on problematic conditions in walls, ceiling, and floors of classrooms in 2022

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Percentage of schools that have at least one classroom with:					
Three or more problems in walls	4.5	72.8	-68.3**	0.000	0.060
Three or more problems in ceilings	0.0	71.6	-71.6**	0.000	0.053
Three or more problems in floors	0.0	62.3	-62.3**	0.000	0.057

Source: Building survey administered in 176 schools in 2022.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table D.4. Impact of rehabilitation on problematic conditions in walls, ceiling, and floors of classrooms in 2022, by timing of the follow-up

	2022 represents the 2-year follow-up		2022 represents a longer-term follow-up		Difference in impacts by subgroup (p-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Percentage of schools that have at least one classroom with:					
Three or more problems in walls	-68.5***	0.000	-68.0***	0.000	0.965
Three or more problems in ceilings	-68.0***	0.000	-78.5**	0.000	0.371
Three or more problems in floors	-31.7***	0.001	-79.5**	0.000	0.000**

Source: Building survey administered in 176 schools in 2022.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

2. Effects on heating systems

In 2022, rehabilitation continued to increase the availability of central heating systems in treatment schools. All treatment schools had functional central heating compared with 72 percent of control schools, and the impact of rehabilitation on classroom temperatures (3.4°C, 6.3°F) was statistically significant as well (Table D.5). The availability of central heating systems was similar among schools where 2022 represents the second-year follow-up and schools that were in their third-, fourth-, or fifth-year follow-up. (Table D.6).

Table D.5. Impact of rehabilitation on the presence of central heating in 2022

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Classrooms					
Have functional central heating	100.0	72.4	29.1**	0.000	0.053
Average measured temperature (med classroom)	18.6	15.1	3.4**	0.000	0.523
Schools					
All classrooms have functional central heating	99.5	66.4	33.0**	0.000	0.057
Indoor gym has central heating	81.2	51.7	29.5**	0.000	0.074
At least one classroom without working lighting in school	4.0	32.4	-28.4**	0.000	0.059

Source: Building survey administered in 176 schools in 2022.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table D.6. Impact of rehabilitation on the presence of central heating in 2022 (subgroup analysis)

	2022 represents the 2-year follow-up		2022 represents a longer-term follow-up		Difference in impacts by subgroup (p-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Classrooms					
Have functional central heating	-31.8**	0.001	-27.5**	0.000	0.704
Average measured temperature (med classroom)	4.3**	0.000	3.0**	0.000	0.241
Schools					
All classrooms have functional central heating	27.5**	0.006	35.1**	0.000	0.535
Indoor gym has central heating	11.4	0.354	37.6**	0.000	0.089^
At least one classroom without working lighting in school	-24.3**	0.004	-30.1**	0.000	0.619

Source: Building survey administered in 176 schools in 2022.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

1. Effects on air quality outcomes

As of 2022, rehabilitation continued to dramatically improve air quality measured as the maximum exposure to $PM_{2.5}$ and PM_{10} levels in rehabilitated schools (with reductions of -45ppm for $PM_{2.5}$ and -63ppm for PM_{10}, compared to the control group). These findings were not driven by the most extreme classrooms with the worst air quality in each school: we also found similarly large improvements in air quality for the median classroom (Table D.7). Improvements in air quality measured as the maximum and median exposure to $PM_{2.5}$ and PM_{10} were similar among schools where 2022 represents the second-year follow-up and schools that were in their third-, fourth-, or fifth-year follow-up (Table D.8).

Table D.7. Impact of rehabilitation on air quality outcomes in 2022

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Average PM 2.5 (max classroom)	14.2	59.2	-45.1**	0.000	11.749
Average PM 10 (max classroom)	32.5	95.5	-63.0**	0.000	15.385
Average PM 2.5 (med classroom)	12.7	40.5	-27.8**	0.000	7.478
Average PM 10 (med classroom)	25.9	85.5	-59.6**	0.000	16.117

Source: Building survey administered in 176 schools in 2022.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table D.8. Impact of rehabilitation on air quality outcomes in 2022

	2022 represents the 2-year follow-up		2022 represents a longer-term follow-up		Difference in impacts by subgroup (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Average PM 2.5 (max classroom)	-41.5*	0.022	-46.0**	0.003	0.850
Average PM 10 (max classroom)	-59.0**	0.009	-63.9**	0.002	0.872
Average PM 2.5 (med classroom)	-24.4**	0.009	-29.2**	0.004	0.727
Average PM 10 (med classroom)	-45.0**	0.007	-65.9**	0.004	0.453

Source: Building survey administered in 176 schools in 2022.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

In 2022, rehabilitation also continued to help schools meet the WHO's interim guidelines for $PM_{2.5}$ and PM_{10} exposure. The improvement in the percentage of schools meeting the WHO long-term exposure goal for $PM_{2.5}$ and PM_{10} were similar among schools where 2022 represents the second follow-up and schools that were in their third, fourth, or fifth follow-up (Table D.9 and Table D.10).

Table D.9. Schools meeting World Health Organization interim air quality targets ($PM_{2.5}$ and PM_{10})

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
School meets WHO's 2021 interim target for air quality ($PM_{2.5}$)	93.8	74.4	19.3**	0.001	0.058
School meets WHO's 2021 interim target for air quality (PM_{10})	93.2	70.4	22.8**	0.000	0.059

Source: Building survey administered in 176 schools in 2022.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table D.10. Impact of rehabilitation on air quality outcomes in 2022

	2022 represents the 2-year follow-up		2022 represents a longer-term follow-up		Difference in impacts by subgroup (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
School meets WHO's 2021 interim target for air quality ($PM_{2.5}$)	23.8**	0.005	16.8**	0.030	0.543
School meets WHO's 2021 interim target for air quality (PM_{10})	23.4**	0.006	22.2**	0.005	0.916

Source: Building survey administered in 176 schools in 2022.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

2. Effects on sanitation outcomes

In 2022 we found no significant differences between the percentage of treatment and schools with a functional flush toilet or running water for handwashing (these results were driven by improvements in control-group schools, rather than any deterioration in the quality of sanitary facilities in rehabilitated schools). The program continued have a significant impact on the cleanliness of toilet facilities in rehabilitated schools (Table D.11). For these outcomes, differences between schools where 2022 represents the second-year follow-up and schools that were in their third-, fourth-, or fifth-year follow-up were not statistically significant (Table D.12).

Table D.11. Impact of rehabilitation on sanitary facilities in 2022

	Treatment (A)	Control (B)	Impact (A-B)	*p*-value	Standard error
Schools without:					
Flushing toilet	13.4	16.2	-2.8	0.606	0.054
Running water for hand washing	2.7	8.7	-6.0	0.131	0.040
Soap near toilets or latrines	5.2	20.9	-15.6**	0.002	0.051
Schools with an odor in restroom facilities	11.6	52.1	-40.5**	0.000	0.067

Source: Building survey administered in 176 schools in 2022.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported).

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

Table D.12. Impact of rehabilitation on sanitary facilities in 2022 (subgroup analysis)

	2022 represents the 2-year follow-up		2022 represents a longer term follow-up		Difference in impacts by subgroup (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Schools without:					
Flushing toilet	8.8	0.513	-8.8	0.087	0.221
Running water for hand washing	0.3	0.624	-9.3	0.117	0.115
Soap near toilets or latrines	-8.3	0.126	-18.9**	0.009	0.235
Schools with an odor in restroom facilities	-31.5**	0.002	-45.2**	0.000	0.312

Source: Building survey administered in 176 schools in 2022.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

This page has been left blank for double-sided copying.

Appendix E

ILEI evaluation impact estimates, by gender

This page has been left blank for double-sided copying.

In this appendix we present treatment-on-the-treated (TOT) impact estimates for the ILEI evaluation, separately for male students and female students. The results shown below represent impact estimates two years after rehabilitation was completed, for each of the student-level outcomes presented in the report. Specifically, for each outcome we conducted a subgroup analysis to assess whether the pattern of impacts was different between male and female students. Broadly speaking, the magnitude and direction of the impact estimates were very similar for male and female students. That said, we did find statistically significant differences between male and female students with respect to the magnitude of the impact estimates for certain outcomes.

A. Effects on school infrastructure

1. Physical condition of the school building

Male and female students both reported substantial improvements in their level of satisfaction with rehabilitated school buildings. Improvements in students' satisfaction with the building was larger for female students compared to male students. Similarly, the difference between treatment and control students who reported that their school needed immediate repairs was larger for female students. The differences between female and male students were statistically significant (Table E.1).

Table E.1. Impacts on student perceptions related to the quality of physical building, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Satisfied or very satisfied with the quality of the building and equipment at their school	60.2**	0.000	50.0**	0.000	0.000**
Reported that school needs immediate repairs or improvements	-75.6**	0.000	-67.7**	0.000	0.000**

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

2. Effects on heating systems

In rehabilitated schools, male and female students both reported large improvements in school heating systems. The differences between treatment and control students who felt that classrooms were too cold in February was larger for female students compared to male students. Moreover, we also found larger impacts on the percentage of female students compared to male students who felt that temperature negatively affected ability to concentrate in winter. The differences between female and male students were statistically significant (Table E.2).

Table E.2. Impact of rehabilitation on student-perceived cold and its effect on learning environment, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Feels classroom is too cold, on average, in February	-33.6**	0.000	-26.8**	0.000	0.002**
Feels temperature negatively affected ability to concentrate in February	-29.6**	0.000	-20.1**	0.000	0.000**

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

3. Effects on air quality outcomes

Both male and female students reported experiencing improvements in air quality in rehabilitated schools. The decrease in the percentage of students who reported that air quality in the classroom in the past month was poor or unhealthy was larger for female compared to male students. However, reports related to air quality affecting student's ability to concentrate and disrupting instruction during cold weather were similar between male and female students (Table E.3).

Table E.3. Impact of rehabilitation on perceived air quality in schools in February, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Air quality in classroom in past month was poor or unhealthy	-28.7**	0.000	-24.0**	0.000	0.026*
Air quality affected student's ability to concentrate in Feb	-20.2**	0.000	-17.0**	0.000	0.149
Air quality disrupted instruction in February	-8.7**	0.000	-10.5**	0.000	0.305

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

4. Effects on lighting and electrical systems

Male and female students also reported similarly large improvements in lighting in rehabilitated schools. As shown in Table E.4, improvements in student reported reading conditions, such as having difficulty reading, having difficulty reading the blackboard, and feeling that lighting negatively affected their ability to concentrate during the winter were larger for female students compared to male students.

Table E.4. Impact of rehabilitation on lighting and its effect on the learning environment, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Ever have difficulty reading because of lighting	-33.5**	0.000	-28.2**	0.000	0.017**
Ever have difficulty reading blackboard because of lighting	-31.1**	0.000	-26.4**	0.000	0.052^
Feels lighting negatively affected ability to concentrate on schoolwork in February	-24.6**	0.000	-20.3**	0.000	0.055^

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

B. Effects on facility use and school safety

1. Use of science labs

Improvements in students' exposure to science laboratories, including receiving more science demonstrations and participating in experiments were similar among female and male students. However, differences in the reported availability of science equipment such as microscopes, heating devices, chemicals, and lab coats were larger for female compared to male students.

Table E.5. Impact of rehabilitation on students' exposure to science laboratories, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
School has a science laboratory	59.0**	0.000	56.7**	0.000	0.452
Students who reported teacher always/sometimes shown demonstrations	34.6**	0.000	30.3**	0.000	0.153
Students always/sometimes conduct own experiments (if teachers did demos)	26.7**	0.000	26.9**	0.000	0.959

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Availability of science equipment (if teachers did demos)					
Equipment not available in science class	-6.6**	0.000	-9.4**	0.000	0.045*
Microscope					
Heating devices	33.1**	0.000	19.9**	0.000	0.565
Chemicals or other materials for experiments	33.8**	0.000	27.0**	0.000	0.053^
Lab coats	41.9**	0.000	25.8**	0.000	0.058^
Protective eyewear	42.6**	0.000	32.6**	0.000	0.000**
Beakers	13.7**	0.000	37.0**	0.000	0.029*
Other science equipment	0.4	0.220	0.1	0.927	0.594
Availability of electronic equipment (if teachers did demos)					
Computer	-3.7	0.339	2.6	0.426	0.066^
Internet access	18.1**	0.000	14.0**	0.000	0.096^
Projection screen	10.1**	0.002	7.5*	0.020	0.446
Television	-1.2	0.640	0.6	0.780	0.208
Other electronic equipment	0.2	0.692	0.6	0.311	0.570

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

2. Use of recreational facilities

Improvement in the usage of indoor recreational facilities was similar among female and male students. While the differences between treatment and control students were statistically significant for both male and female students, results for both genders are statistically indistinguishable from one another (Table E.6).

Table E.6. Impact of rehabilitation on use of recreational school facilities

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Availability of indoor gym	7.5	0.076	6.6	0.103	0.641
Student reported using at least once in an average week					
Indoor gym (if available)	7.0**	0.006	5.1*	0.041	0.241
Outdoor recreation area (if available)	4.7	0.299	1.0	0.822	0.260

Source: Student surveys completed by 7,978 students, interviewed at one-year follow-up.

Notes: Columns A and B present two-stage least-squares (2SLS) regression-adjusted group means. 2SLS estimates the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment and control group are significant at the 10/5/1 percent levels.

3. School safety

Male and female students both reported large improvements in safety in rehabilitated schools. However, the magnitude of these improvements was larger for female students compared to male students (Table E.7).

Table E.7. Impact of rehabilitation on perceived safety, by gender

	Female students		Male students		Difference in impacts by gender (*p*-value)
	Difference (T and C)	*p*-value	Difference (T and C)	*p*-value	
Agrees that the school is safe and healthy	48.0**	0.000	36.9**	0.000	0.000**
Feels very safe in the classroom	44.9**	0.000	33.1**	0.000	0.000**
Feels very safe using stairwells	49.7**	0.000	38.7**	0.000	0.000**

Source: Student surveys completed by 8,085 students, interviewed at two-year follow-up.

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

C. Effects on learning outcomes

Differences between study groups in test scores in language and math were very similar among female and male students. However, the pattern of science learning impacts is substantially different: for female students the impact is positive but not statistically significant, whereas for male student the impact on science learning is negative and statistically significant (Table E.8). Note, however, that this pattern of gender-specific differences in science learning may have been related to the COVID-19 pandemic: prior to the pandemic rehabilitation had similar effects on science learning for males and females (for example, among males the activity had an impact on science learning of -0.05 standard deviations; p-value of 0.66). However, during the pandemic rehabilitation appears to have had a more pronounced negative effect on science learning among males (with an impact estimate of -0.22 standard deviations; p-value of 0.10).

Table E.8. Impact on rehabilitation on student test scores across grades 10, 11, and 12, by gender

	Female students		Male students		Difference in impacts by gender (p-value)
	Difference (T and C)	p-value	Difference (T and C)	p-value	
Language	0.03	0.714	0.00	0.972	0.787
Math	0.06	0.465	-0.02	0.835	0.382
Science	0.08	0.361	-0.21	0.029*	0.012*

Source: Education Management Information System (EMIS) administrative data from 150 schools and 3,695 students..

Notes: The differences between treatment (T) and control (C) using 2SLS estimates show the impact of school rehabilitation on schools that received the program. Regressions included controls for the probability that each school could be selected into the treatment group (not reported). Standard errors were clustered at school-level.

^/*/** indicates that differences between treatment (T) and control (C) group are significant at the 10/5/1 percent levels.

Appendix F

Responses to stakeholder comments

This page has been left blank for double-sided copying.

Table F.1. Responses for stakeholder comments

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	xviii, Figure ES.4; Section IV Figures	There is no x-axis title. While I think that it can be assumed, I would suggest adding something for completeness. This is true for those in Section IV as well.	Figures ES.4, IV.1, and IV.2 were updated adding an x-axis title.
MCC Economist	General	The Georgia school year months are not stated explicitly, particularly within the TEE Activity when noting data collection took place each of the years in September. I believe their school year starts in September, so this timing seems interesting and worth noting why this timing was chosen for the data collection. This would seem to be when they would just be ramping up and not have gotten into the real teaching/learning yet.	We have added a discussion of the Georgian school year and how the timing of the TEE teacher surveys relates to the start of the school year (see section II.B.3).
MCC Economist	Pg. 20, Table II.6	The table could benefit from a clear visual distinction (potentially labeling or color coding) between what was covered in the interim report and the focus of this Final report, as well as whether there is any overlap in the two. While it has been described above, the table provides an opportunity to summarize and make the distinction clearer to readers and not require them to figure it out on their own. The executive summary notes the last two questions as those of interest, but it seems that other pieces, following up from previous results may be included here too, and the narrative in this actual section of the main text seems less clear in describing the focus of the final evaluation.	We have labelled the table to show which research questions and analyses were completed for the final report.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	Overarching	Baselines: A notable gap to me in the report is not sufficiently addressing the topic of baseline results. In Table II.6 (pg. 20) the two main research questions are outlined. These seem difficult to answer without a baseline - i.e., pre-training data on teacher practices. If it doesn't exist then this should be reflected on as whether this is a potential weakness or not, and if it is then how do the analysts try to overcome this weakness? Strangely, there is a brief mention of baselines for the first and only time on pg. 22 in Table II.7 for Cohort 2. This further begs many questions that are not addressed in the report: Was baseline data was collected? Was it collected for both Cohorts 1 and 2? If yes, then what was found? Why is this not used in the analysis of the final report? If not collected, then what does a lack of a baseline mean for interpreting the results?	Baseline data for Cohort 2 was collected and analyzed in this evaluation's interim report (Nichols-Barrer et al. 2019). The interim report presented a pre-post analysis of training outcomes for Cohort 2 (comparing outcomes at baseline to outcomes one-month after training) and a matched-comparison group analysis comparing the baseline outcomes of Cohort 2 to the outcomes of a matched set of Cohort 1 teachers shortly after they completed the training sequence. This final report focuses on longer-term trends in teacher outcomes after training was completed, and does not present these analyses of baseline data because it would be duplicative of the interim evaluation report. That said, high-level findings from the interim report (including these two analyses of baseline data for Cohort 2 teachers) are summarized briefly in the opening pages of Chapter IV.
MCC Economist	several pages; Follow-up period	Notable gap in report: The description of data collection and the results are comparing a post-training of one year for Cohort 2 and two years for Cohort 1. I did not see this explicitly stated, or addressed within the description of what this could mean for interpretation of the results. We don't know where C2 was at 1 year after training. These are not exact comparisons because data appears not to have been collected in Sept. 2018 on C1 to understand the 1 year after situation. If that is incorrect, then the language above needs to be adjusted to more clearly describe and report on post 1-year results for C1. This difference will be picked up from the reader and should be clearly addressed.	We understand this concern and have added clarifying details about these follow-up periods in the report's methods section (see section II.B.1). In addition, note that 1-year follow-up data for Cohort 1 was collected and analyzed in this evaluation's interim report (Nichols-Barrer et al. 2019). Specifically, the interim report presented a 12-month follow-up trend analysis for Cohort 1 (comparing outcomes one-month after training to outcomes one year later). This final report focuses on longer-term trends in teacher outcomes after training was completed, and does not present these analyses of Cohort 1's one-year follow-up data because it would be duplicative of the interim report. That said, high-level findings from the interim report are summarized briefly in the opening pages of Chapter IV.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	Pg. 60 (and beyond); Follow-up period	Notable Gap: It seems problematic that the information at the end of the page 60 is not provided until this point in time and that it is unclear how this is incorporated into the analysis that has been conducted and the results presented in this section. Specifically, it states that "Cohort 1 teachers were given an opportunity to make up for any training modules that they missed and sit in on training events attended by Cohort 2 teachers." It follows by stating that this option actually led to a notable increase with 'Cohort 1 teachers who attended any TEE training modules increased from 64 to 82 percent after the second round of training." This means that the simplified dichotomy that has been set out from the beginning is actually not a clear division into two groups. This raises a lot of questions to the reader and doubt on interpreting the findings, and this topic is not fully addressed within the report. At a minimum, I think that the report needs to follow-up by indicating (1) how many people fall into the category of being both in Cohort 1 and Cohort 2; (2) how are Cohort 1 and Cohort 2 actually defined, as the 64% to 82% says that this changed whether they had taken any training, so why are those individuals in Cohort 1 , if they didn't actually take training until the time with Cohort 2?; (3) How many individuals fall into this category of individuals, "assigned" to Cohort 1 but actually were in Cohort 2? ; (4) At what point would those in Cohort 1 become Cohort 2? Did that change in grouping happen for the analysis at all? - for example, were there cases where Cohort 1 teachers actually took all their training with Cohort 2? (5) What is the average time between completing the last training and the data collection for "two years after"? Overall, this questions the accuracy of the findings and will need to be addressed.	To the extent these cohort-crossover patterns affect teacher outcomes, the issue makes the follow-up period for Cohort 1 (two years following the end of their assigned training year) more comparable to the follow-up period for Cohort 2 (one year following the end of their assigned training year). The analyses presented in this report are descriptive in nature, and do not support causal inferences about the impacts of training on teachers. In contrast, in our interim evaluation report (Nichols-Barrer et al. 2019) we presented the results of a matched comparison group analysis that was designed to estimate the effects of training participation on teachers; that analysis was conducted at a point in time (shortly after Cohort 1 completed training, and before Cohort 2 could begin training) when these types of cohort-crossover issues could not bias the impact findings. For the trend analyses in this final report, however, the evaluation is only tracking changes in teacher outcomes over time (among all teachers who received an invitation to participate in training in Cohort 1 or Cohort 2). The study is not designed to separate out the effect of training on Cohort 1 teachers who completed training in their assigned year versus the effect among Cohort 1 of 'sitting in' on training modules in a later training year. We have made clarifying edits section II.B.3 to explain these points in more detail.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	Pg. 21, and more general	The report states that the performance evaluation did not include student learning assessment or student exams. Given what we know about the reported and observed changes in teacher practices by 2019, this decision seems to make sense as we don't seem to be further along the logic chain in reaching certain outcomes that would be measured by test scores. However, at the same time, this was a nationwide program, and I imagine that there are national test scores. I cannot remember how frequently these are collected, and I am sure there were disruptions during COVID, but I think it would be worth noting that national data is available, but there was a decision not to review that given what I note above or that the MPR, or if MPR thinks that it is worth getting that data to review then consider that. No misperception that this would be as rigorous, but if disaggregated data were available by school and you have teachers in those same schools, I just wonder if there would be anything useful from that exercise. The gap is a bit more visible as the other activity within this Project, which is not being included in this evaluation and the report, is on student assessments.	The National Assessment and Examinations Center (NAEC) did not conduct universal, nationwide learning assessments for secondary-level students in 2020, 2021, or 2022 so the suggested analyses are not possible for the TEE evaluation. We have noted this information following Table II.6.
MCC Economist	Student learning	Related to the comment directly above, the School Rehab findings on 'effects of learning outcomes' (starting on pg. 44) do not mention that teacher training occurred in all of these schools and that was the intervention that was expected to have the larger impact on student learning. While these outcomes may not have been realized as of 2019, it seems like a notable gap to not be explicit about the TEE Activity, their relationship, and that teacher training took place in all of the control and treatment schools, and by the year of Rehab data collection the project logic assumes that the learning aspects were being realized to at least some degree. Since there was no overlap with this group in the TEE sample, this cannot be explored further, but the TEE findings do suggest greater results in science and that I believe may jive with the findings seen for Rehab. I would recommend greater reflection on this, and being more explicit about these topics, as it will be on readers' mind.	Because the school rehabilitation study is a randomized controlled trial, and the rollout of the TEE activity was nationwide (giving no preference to rehabilitated or non-rehabilitated schools), treatment and control schools both had the same level of exposure to TEE trainings over the course of the evaluation. As a result, any gains in student learning produced by the TEE activity would have been present in both treatment and control schools and would not affect the school rehabilitation impact estimates in this report. The most accurate way to interpret the evaluation's findings is to treat the TEE evaluation and its findings as being completely separate from the school rehabilitation study and its findings. We have made clarifying edits to section II.A.3 to emphasize this point.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	pg. 22, and more	Linkages and decisions around sampling between the two main activities: Rehab and Teacher Training. (1) top of pg. 22 - it is briefly noted "the sample frame also excluded all of the schools participating in the evaluation of school rehabilitation". It seems like a gap to not indicate why this decision was made, as well as explicitly stating whether this extended to both the treatment and the control schools.	We have clarified the text here to explain that the decision to exclude the treatment and control schools in the school rehabilitation study sample from the TEE study sample was made to limit burden on respondents and ensure that the evaluation could assess the assumed links between each activity and the project's outcomes of interest on a separate basis.
MCC Economist	Table IV.4	I am wondering whether this is the right table to highlight/include as this focuses on the differences, but does not actually indicate who was performing better in the two time periods. One could have reduced more from period 1 to period 2, but remain much higher than the other group of teachers, but that wouldn't be captured here. I am concerned that the info provided alone does not provide a complete picture of the situation, and this is leading to some differences as being significant when small and large ones not being significant, and not knowing what they are being compared to determine that significance. RELATED: It is unclear what significance is being reported. My assumption was that the table was providing significance on results by differences between the two groups of teachers' changes across time, but a few of the markings made me think that it could be reporting significance by X type of teacher between two periods in time. I would suggest making this explicit in a footnote below the table - see note below of changes needed to current significance footnote. It may be useful to go back and review that the same calculation was done for all practices. For example, see 'Work with struggling students: Every day?' the difference is only 0.01 yet it shows as significant. I would think that it wouldn't be if you were comparing -0.05 and -0.06, however the values for each are not indicated here. As comparison, 'Use ICT in instruction: Every week?'. There is a difference of 7pp, yet not significant. That seems surprising.	We have updated the footnote in Table IV.4 to explain that we performed significance tests separately for each group of teachers (comparing the 2017 mean for a given group to the 2019 mean for that group). We did not conduct significance tests estimating if the trend-lines for the two groups were equivalent. In addition, we did not add columns showing the 2017 and 2019 means for the two groups because the amount of detail in the table (with 4 additional data columns) would likely become overwhelming to readers. The mean outcomes of senior teachers and practitioner teachers were also discussed in the evaluation's interim report (Nichols-Barrer et al. 2019), so we did not include a comparison of the two groups here as it would be duplicative of the prior report.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	pg. 55, 63, 65, and more	Notable gap: (1) Report requires clear definition and consistent use of specific terms to explain junior vs. senior and practitioners vs. senior. There appears to be a distinction by type of teacher: Practitioner and Senior. However, the term junior is also used and compared to senior which gives the impression of an age or years of experience aspect. There are many places where this creates confusion on what is being compared in the analysis and reporting, particularly where Cohort 1 is associated with more "senior" teachers and Cohort 2 is more associated with Cohort 2 teachers, and in some places cohort is not noted but comparisons between junior/practitioner and senior teachers is noted, and it is unclear whether further disaggregation was completed for analysis that is not shown in a table or if this is simply a reference to Cohorts 1 and 2, respectively. (2) Additionally, given this confusion but continuous reference, it needs to be explicitly reported early in the report on the percentage breakdown between junior/senior and/or practitioner/senior for each of the Cohorts. This is referenced in many places but not provided in percentage terms. (3) Lastly, this begs the question of how much of the results we are seeing are attributed to the training provided or the experience/level of teacher that is taking the training - with considerations on the timing aspect of course. This is where the baseline aspect comes back into play.	We have removed references to the term 'junior teacher' throughout the report, clarified throughout the report that our cross-cohort trend analyses only compare the aggregate Cohort 1 teacher sample to the aggregate Cohort 2 teacher sample, and presented a more detailed explanation of these teacher-qualification terms in section II.B.2. This section also now summarizes data from the interim report on the percentage of practitioner-level teachers in each cohort.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	Overall/General	A notable gap seems to be any recognition about what has happened since the training programs were implemented. While I see that sustainability is not a specific research question, I would think that a final evaluation would at least reflect on what has happened in the last 3-4 years. Are new teachers being trained on the same curriculum? What was done during COVID? Were the trainings also moved to remote platforms? - I believe that was a step they were going to take after MCC left. What kind of follow up training has been happening to reinforce the principles and practices outlined in the initial training? I had thought that plans had been laid out on how this would be continued. Are they doing these aspects? The CBA model is considers a 20-year time horizon, so this is particularly important for those assumptions as well - but clearly useful for all of MCC. Additionally, this is relevant to an earlier comment on whether there are no student outcomes differences between control and treatment groups for Rehab because any gains would have come from the TEE Activity and that was implemented nationwide.	While we agree these questions about post-Compact training initiatives are interesting and potentially very useful, they are not part of the evaluation design established for the TEE activity. The quantitative and qualitative data collection activities commissioned by MCC in the post-Compact period focused on the school rehabilitation study rather than the TEE study.
MCC Economist	Figure IV.2	For this table and the one above, it would be helpful to see if you could insert some kind of significance notification - especially since there is no table provided that outlines those findings. This would make it easier for the reader to see the results.	Because the follow-up period differs between Cohort 1 and Cohort 2, our analysis plan did not include tests for statistical significance when comparing the follow-up means of the two samples (as noted in the methods chapter in section II.B.1). These two figures are only meant to be descriptive in nature, showing the magnitude of the differences in means for the two cohorts.
MCC Economist	Most tables in Section IV	The notes under the tables stated "indicates that differences between treatment and control group are significant . . . ." . There is no treatment and control group for the TEE Activity, so all these footnotes for the tables need to be adjusted.	We have made this change.
MCC Economist	Table IV.4	For clarity, suggest including Cohort 1 in title.	We have made this change.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	Table IV.5, pg. 63-64	Suggest moving this table up, as it is referenced on the first line of page 64 and before Figures IV.2 and IV.3, yet is placed on page 66. This would focus on presenting and discussing the findings for Cohort 2, and then talking about comparing Cohort 1 and 2, which is captured in those two figures. Otherwise, change ordering of findings and narrative to describe the figures first and then the table.	We have moved up Table IV.5 to align with the first paragraph where the table is mentioned.
MCC Economist	pg. 63, mid-description of C2 vs. C1 finding	Language in sentence defines comparison calculation for 2019 differences but then this follows with the results of one month after they completed their trainings which would have been September 2017 for C1 and September 2018 for C2. Please adjust to avoid confusion and sync calculations with what is shown.	We updated this paragraph to clarify the reference years for each cohort.
MCC Economist	xii-xiii	On increased operating costs due to school rehabilitation: At the time of Closeout CBA model completion, I did not have data on differential operating costs (mostly related to additional heating costs in with-Project schools) - did Mathematica collect quantitative data on these costs and, if so, could it be shared? Further, while sustainability of the benefits of the investment are supported by information about additional student revenues, the incremental O&M costs would not be offset by student fees in a CBA model if the school fees were the same across schools and students were transferring these fees to a different school (no net change in fees for beneficiaries or no new student enrollees/students who were not transfers). It might also be helpful for context to explain how increased enrollment increase school revenues. Are students paying fees for secondary or do schools receive a budget from the government based on the number of enrollees?	All of the study's information about utility and operating costs in treatment and control schools is shown in Table III.17. School revenues are determined by the Georgian government's per-pupil funding allocation (families do not pay school fees directly). Each school's funding allotment is calculated using a formula based on enrollment levels in the school at the start of each school year. This RCT was designed to estimate the impacts of rehabilitation at the school-level (not nationwide); in the context of this study design, attracting a transfer student would represent a net increase in a rehabilitated school's revenues, even though it also represents a net loss in revenue for the school that student used to attend.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	xiii	The logic in the M&E Plan and in the CBA model for ILEI assume more time on task as a result of rehabilitation, but teachers were not found to spend more time in classroom instruction in with-Project schools. There does seem to be some nuance around classroom focus during instruction time, even though instruction time did not significantly change as a result of the Project. Was classroom observation able to measure increased "focus" time in a quantitative way, i.e., to measure decreases in student distraction during instruction time, which would reflect a higher quality of instruction and/or learning?	We did not conduct classroom observations for the school rehabilitation study (classroom observations were only performed for a subset of teachers in the TEE study). However, qualitative teacher interviews strongly suggest that teachers in rehabilitated schools may have only been reporting the amount of time they attempted to provide classroom instruction, without accounting for interruptions during scheduled class periods. Based on qualitative interviews with teachers, as well as survey findings about how improved infrastructure affected the learning environment, it does seem plausible that a team of classroom observers could have detected increases in the amount of focused instructional time in rehabilitated schools.
MCC Economist	xv (and more broadly)	The ILEI RCT measured a positive (but insignificant) change in test scores in the pre-COVID cohorts, which could be a driving benefit stream as the evaluation found little in the way of increases in attendance or decreases in dropout rates that were assumed in the Closeout CBA model. However, the ILEI Activity beneficiaries are a subset of the TEE beneficiaries, where logic and CBA models for TEE also assume changes in test scores that were not measured in the TEE evaluation. It might be useful to include a discussion on how the evaluation does not make it possible to separate changes in test scores due to improved teaching as a result of teacher training from changes in test scores due to improved school environments/rehabilitated schools. As the evaluation states, the literature on school rehabilitation is weak on its impacts on learning quality, and we were unable to learn through this evaluation whether there are complementarities between school rehabilitation and teacher training or if there are issues of double counting of benefits if we assume learning increases are additive for ILEI and TEE.	Because the school rehabilitation study is a randomized controlled trial, and the rollout of the TEE activity was nationwide (giving no preference to rehabilitated or non-rehabilitated schools), treatment and control schools both had the same level of exposure to TEE trainings over the course of the evaluation. As a result, any gains in student learning produced by the TEE activity would have been present in both treatment and control schools and would not affect the school rehabilitation impact estimates in this report. The most accurate way to interpret the evaluation's findings is to treat the TEE evaluation and its findings as being completely separate from the school rehabilitation study and its findings. We have made clarifying edits to section II.A.3 to emphasize this point.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	xv	Discussions with the Infra lead around low levels of O&M budget commitment from GoG at CED led to a reduction in the assumption of beneficiary cohorts (from 20 to 10), and a lowered assumed annual contribution to/cost of school maintenance over a 10-year period (versus a higher amount over a 20-year period). Do the findings of maintained improved infrastructure in with-Project schools after five years suggest that there is merit in extending benefits of infrastructure improvement to the standard 20 cohorts considered in education CBA models? Further, is there any updated data on maintenance costs before and after the Project?	All of the information collected by the evaluation about utility and operating costs in treatment and control schools is shown in Table III.17. Because the evaluation's follow-up period does not extend beyond 5 years, the data from this study cannot definitively show whether it would be more accurate to model the project's benefits over 20 cohorts rather than 10 cohorts. That said, on average we did not find any evidence of meaningful deteriorations in infrastructure quality 1-5 years after these investments were made. Based on the data from this evaluation, it appears that any decreases in the assumed benefit stream should only begin after Year 5.
MCC Economist	4	For the mention of ERR calculations, what was the reference ERR for this? IM, EIF, or Closeout? The increase in graduation and transition rates is correct for the Closeout model, but a learning quality parameter was also included based on discussions around the potential for science lab construction to increase test scores (which the evaluation found largely didn't happen).	Thank you for clarifying that this additional parameter was included in the Closeout ERR model. We have added a clarifying note about this in section I.B.1.
MCC Economist	4	The literature review indicates that improvements in physical infrastructure of schools can increase enrollment and attendance, but impacts depend on the baseline condition of the schools. Control school data indicate that pre-rehabilitation, schools were unlikely to be functioning well - what would be Mathematica's assessment of initial conditions and how they compare to those presented in the available literature?	We agree that prior to rehabilitation there were a wide range of serious infrastructure problems in the schools selected for rehabilitation, and have added a clarifying note about this in section I.B.1.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	10	Was there a high level of correlation between the schools that had existing rehabilitation designs and the other school-level characteristics? Similarly, were schools excluded from random assignment due to needs exceeding rehabilitation correlated with other school-level characteristics from the stratification process?	Yes, the schools that had pre-existing rehabilitation designs prior to random assignment were selected in a broadly similar manner to the schools that were randomly assigned later. We also did not see evidence that the treatment schools excluded from rehabilitation due to feasibility constraints differed systematically from other schools in the treatment group on the school-level characteristics used for the stratification process. These exclusions tended to occur for idiosyncratic reasons, such as a school's foundation showing levels of damage that would have made it more cost-effective to demolish and rebuild the school instead of rehabilitating the site.
MCC Economist	17	Could the assumption that random assignment to the treatment group had no effect on treatment schools that were not rehabilitated fail if enrollment increased at treatment schools in response to selection for treatment, but where rehabilitation was ultimately not done? If so, how would this affect results/what are the implications for parameter estimates?	In the context of an RCT, intention-to-treat (ITT) impact estimates, which only measure the impact of being assigned to the treatment group, are not subject to this assumption. As we show in Appendix A, the pattern of large positive effects on school infrastructure and learning-environment outcomes that we present in the main report (using a treatment-on-the-treated [TOT] model) does not change when the analysis uses an ITT model instead. As a result, even if it were true that assignment to the treatment group had an effect of some kind on non-rehabilitated schools, this (hypothetical) violation of the assumptions in the TOT model would not change any of the evaluation's substantive conclusions.
MCC Economist	18	Can you briefly explain why you think room temperature didn't come out as a constructed index for the impact study? The data indicate significant qualitative and quantitative differences in room temperature/comfort, but this didn't seem to come out of the PCA.	Additional details about the principal-component analysis can be found in the evaluation's interim report (Nichols-Barrer et al. 2019). The evaluation only constructed indices for broad school infrastructure outcomes that were measured in multiple different ways in the infrastructure assessment. Since classroom temperature was theorized (ex ante) to be an important predictor of the quality of the classroom learning environment, and classroom temperatures were only measured using a single methodology in each classroom, the evaluation team elected to examine this outcome separately rather than including classroom temperatures as one component of a broader index.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	23	The difference between the 91 rehabilitated schools and the 88 randomly assigned schools might need to be explained earlier on and in-text (a footnote citing the Star Report could remain).	We have noted this issue earlier in the report, in section II.A.1
MCC Economist	35	The evaluation finds girls and boys experienced similar improvements in comfort using sanitary facilities. Did any qualitative questions probe on whether male or female students were more likely to attend school because of improved restroom facilities?	The evaluation conducted mixed-gender student focus groups, and these focus groups provided an opportunity for students to discuss their level of comfort with sanitary facilities. Transcripts from the focus groups did not suggest that access to improved sanitary facilities had an effect on the likelihood of students to attend school (either for males or females). Since these were mixed-gender focus groups, the evaluation team determined that it would not be appropriate to include an extensive set of follow-up questions asking about differences in how male students versus female students responded to these facility upgrades. Instead, we tested for gender differences using quantitative student survey data. These results appear in several places throughout Chapter IV, and are summarized comprehensively in Appendix D.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC Economist	38	Rehabilitation was not found to increase teacher instruction time. Was there any qualitative evidence suggesting what was done with the time saved looking for suitable space to hold class/lighting wood fires? Or were there any quantitative assessments of time savings for teachers who no longer had to do these tasks?	We did find qualitative evidence that rehabilitation increased the amount of time teachers are able to spend on classroom instruction (as discussed in Chapter III of the report). Our view is that teachers responding to our quantitative surveys about time-use simply reported the number of minutes they attempted to spend on instruction each day, and did not reduce these estimates to account for interruptions related to school infrastructure during scheduled instruction time. In qualitative interviews, for example, teachers in rehabilitated schools reported that instructional time increased following improvements to heating and electrical systems. One teacher from a rehabilitated school said that they used to only be able to teach for 30 minutes out of each 45-minute lesson during bad weather because managing the wood heating system was time consuming. In another school, a teacher mentioned that prior to the rehabilitation, class was disrupted by wind blowing smoke from the wood stove heating system into the classroom and causing irritation to the eyes; the class would have to move to the hallway to find relief from the abrasive smoke. Additionally, several teachers report that improved access to electricity helps them run more efficient lessons using technology (such as projectors or audio/video systems).
MCC Economist	41	Were there any "best practice" examples of schools who strategically scheduled different types of science classes to ensure increased availability of the lab? Was there any evidence that time saved due to improved general classroom conditions was used by some teachers to set up a lab, move students to a lab from the classroom, etc.?	The interview and focus group guides used in this evaluation were not designed to elicit extensive information about lab scheduling practices. That said, a few school directors did mention that they use a coordinated schedule across science teachers, to avoid situations where multiple teachers expect to use the lab at the same time. Students and teachers also mentioned that they limit their lab use to the lessons that actually require the lab facility, to avoid occupying the lab when other teachers might need it.
MCC Economist	52	More on operating costs (Table III.17): do we have any evidence of increases in annual costs? February is probably the most expensive month for heating and electricity (perhaps not water), so constructing an annual incremental operations cost would not simply require multiplying the differential cost for February by 12.	We have added an additional set of impact estimates to Table III.17, showing results for an additional survey item that asked directors to report how often they were unable to pay utility costs. We did not collect annualized cost estimates from directors as part of this evaluation.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC M&E	xvii	The topline findings are confusing as written, please explain that self reported confidence in pedagogical practice may not translate to actual knowledge or changes in teaching practices. This is especially important because the evaluation question asks about knowledge, not confidence. In addition, the bolded sentence "there were large improvements in the use of teaching practices related to students' critical thinking..." should be clarified. These are practices that MCC, MCA-G, and TPDC *thought* would improve critical thinking.	We have made clarifying edits addressing these two issues.
MCC M&E	3	Please re do the Program Logic Graphic - it's fuzzy and you cant read anything. It would also be great to update it to show which causal pathways you believe were supported by the evaluation findings. Maybe also note that in implementation the TEE and School Rehabilitation activities were implemented separately and it was hard to evaluation the overlapping effects as visualized in the original program logic.	We have updated the logic model graphic and highlighted the causal pathways in the model that were assessed in the independent evaluation.
MCC M&E	4	It would be useful to point to some of the critiques of the original program logic here. We've discussed that years of schooling and attainment were not the largest benefit streams but those were what were modeled in MCC's CBA	We have added a discussion of these points.
MCC M&E	4	When describing air quality and health outcomes outside of the program logic, it's worthwhile to reemphasize that these do not contribute to MCC's stated objective. In the context of this evaluation health outcomes that have not led to learning improvements are purely secondary	We have added a reminder that health outcomes were not part of the program logic.
MCC M&E	5	rephrase "importantly we did not find any rigorous prior studies" "We did not find any prior studies that assessed the impacts with rigor"	We have made this change.
MCC M&E	15	It's worth explaining a little bit how rounds of data collection were shifted and added due to implementation delays and then due to COVID.	We have added a more detailed discussion of these decisions in section II.A.3.
MCC M&E	26	I wonder if it's worth while to point out that an upside to the delays in endline data collection is that we are able to observe the infrastructure several years after implementation	We have added additional discussion of this issue in section II.A.4.

Reviewer Institution	Page Number	Comment	Response from the evaluation authors
MCC M&E	39	Here is a good place to have a more comprehensive discussion about that intermediate outcome as part of the program logic and that there are different ways to measure class time	We have added a more detailed discussion of this topic in the first paragraph of section III.C.1.
MCC M&E	50	Is it worth reiterating that students in Georgia have the ability to enroll in a different school, and choose the newer one?	We have added a reminder of this issue in section III.F.1.
MCC M&E	58	For TEE I would again reference back to the program logic, especially because we were not able to measure student learning as it related directly to this intervention	We have added an additional discussion of the program logic for the TEE activity (and the absence of student learning outcome data in this study) in the opening section of Chapter IV.
MCC M&E	multiple	For schools that were rehabilitated earlier on in the Compact, is it possible to investigate if school rehabilitation had longer-term (three-year or four-year) effects on enrollment levels?	Yes, we have estimated 3-year and 4-year impacts on enrollment levels in the subset of schools where we can observe impacts more than two years after rehabilitation was completed. As expected, the data in these schools show a continuing trend of larger enrollments beginning in grade 1 (2-year impacts) and then continuing to grades 1 and 2 (3-year impacts) and grades 1-3 (4-year impacts). We have added a discussion of these results in section III.F.1.
MCC Human and Community Development	multiple	This is a really interesting report, congrats to Mathematica. My only question/comment is whether they broke down findings between Georgian-language and the minority Russian- , Azerbaijani- and Armenian-language speaking teachers? It would be interesting to see whether the training given to the Georgian speakers has the same impact/results as that provided to the minority language speaking teachers.	For the TEE evaluation, our sample is limited to Georgian-language trainees because the activity limited the first two cohorts of trainees to Georgian-language teachers. Minority-language teachers received the TEE training sequence in a third cohort whose training sequence ended in late 2019. The timing of the training sequence for these Cohort 3 teachers did not align with the timeframe of this evaluation, so they could not be included in the analysis.
Georgia Teacher Professional Development Center	multiple	The report requires clear definition and consistent use of the term practitioner teacher vs. senior teacher, which pertain to whether teachers have passed a set of qualification exam requirements and is not related to teachers' years of experience (as implied by the term 'junior' teacher).	We have removed references to the term 'junior teacher' throughout the report, and presented a more detailed explanation of these teacher-qualification terms in section II.B.2. This section also now summarizes data from the interim report on the percentage of practitioner-level teachers in each cohort.

Mathematica Inc.

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Operating in Tanzania, Uganda, Kenya, Mozambique, and the United Kingdom

mathematica.org website