CENTER FOR GLOBAL DEVELOPMENT

Improving School Leadership in Rwanda

Simeon Lauterbach, Lee Crawfurd, Jocelyne C. Kirezi, Aimable Nsabimana, Jef Peeraer

Abstract

This study investigates the short-term impacts of a school leadership professional development program implemented in 525 randomly selected schools across Rwanda from 2018 to 2019. The program aimed to enhance the skills of school headteachers in leadership, management, and teacher support. Although no significant average treatment effects are observed one to two years after the intervention, an increase in test scores is identified in public primary schools compared to government-aided schools by at least 0.11 standard deviations. This disparity may be attributed to the potentially weaker school management and resources in public primary schools at the outset, as well as the time constraints and ownership structure faced by headteachers in government aided schools. Despite the modest effect, the program shows potential for cost-effective improvement in student learning, especially considering that typically only one headteacher per school is trained. Further research should focus on optimizing the design of school leadership professional development programs and exploring the underlying mechanisms necessary to enhance their overall effectiveness.

KEY WORDS
School leadership, professional development, school management

JEL CODES
I22, I28, O15

WORKING PAPER 691 · APRIL 2024

Improving School Leadership in Rwanda

Simeon Lauterbach
Geneva Graduate Institute (simeon.lauterbach@graduateinstitute.ch)

Lee Crawfurd
Center for Global Development (lcrawfurd@cgdev.org)

Jocelyne C. Kirezi
VVOB - education for development (Jocelyne.Cyiza.KIREZI@vvob.org)

Aimable Nsabimana
UNU-WIDER (aimable.nsabimana@wider.unu.edu)

Jef Peeraer
VVOB - education for development (jef.peeraer@vvob.org)

This paper benefited from constructive comments and suggestions by Carlotta Nani, Cedric Tille, Eeshani Kandpal, Jacobus Cilliers, Lore Vandewalle, Loran Pieck, Marko Mlikota, Nathan Sussman, Paula Gonzalez, and Rui Esteves. We wish to express our profound gratitude to the National Examination and School Inspection Authority (NESA) and the Rwanda Basic Education Board (REB) for granting access to standardized students' test score datasets and to the Rwanda Ministry of Education (MINEDUC) for sharing official documents related to the education sector in the country. All errors and views expressed in this paper are those of the authors. They do not necessarily represent the views of the GGI, CGD, UNU-WIDER, or VVOB.

Simeon Lauterbach, Lee Crawfurd, Jocelyne C. Kirezi, Aimable Nsabimana, Jef Peeraer. 2024. "Improving School Leadership in Rwanda." CGD Working Paper 691. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/improving-school-leadership-rwanda

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036
202.416.4000

1 Abbey Gardens
Great College Street
London
SW1P 3SE

www.cgdev.org

Center for Global Development. 2024.

1 Introduction

The quality of school management is an important factor in explaining the current worldwide learning crisis (Muralidharan and Singh, 2020). Although many countries have pledged to provide quality inclusive education as part of the United Nations Sustainable Development Goal (SDG 4), progress towards achieving this goal by 2030 has been slow. Estimates suggest that almost two thirds of 10-year-olds lack the ability to read or comprehend basic texts, making it imperative to take swift and effective actions to avoid the global learning crisis becoming a lasting disaster for future generations (UNESCO, 2020). Unprepared students, poor teaching quality, and poor school management are key factors preventing educational progress (World Bank, 2018).

School management, particularly in low- and middle-income countries, has been identified as an area of concern (Bloom et al., 2015; Lemos et al., 2021). School leaders' management practices can explain a significant proportion of student achievement and are therefore an important component in the daily life of a student (Crawfurd, 2017; Branch et al., 2012). They are responsible for the day-to-day operations of a school, but also for selecting and supporting teachers, monitoring the budget, maintaining the school's facilities, and fostering positive relationships with the wider community (Grissom et al., 2021; Miller, 2013).

Given the evidence on the importance of management practices, two questions remain: Can high-quality training programs improve student learning outcomes, and, if so, can they be cost effective when delivered at scale (Anand et al., 2023; Bloom et al., 2015; Lemos et al., 2021)? We address these questions by analyzing the impact of a large School Leadership Professional Development Program (SLPDP) implemented in Rwanda in 2018 and 2019. The program was designed to strengthen leadership, management, and teacher support skills, with the overall objective of improving the learning environment and student achievement.

The SLPDP was rolled out across three randomly selected cohorts in six districts. The first cohort received training in 2018, followed by the second cohort in 2019. The third cohort was scheduled to receive training in 2020 but did not have the opportunity due to the COVID-19 pandemic. The head teachers were invited to participate in the training program and the selection of headteachers to be contacted was randomly assigned among the three cohorts.

We leverage this randomization to estimate the effect of the program on students' test scores at the Primary Leaving Examinations (PLE). Since not all invited headteachers participated in the SLPDP, we measure the intent-to-treat effect of the program. On average, we find that the SLPDP had no effect on student test scores measured during the school year the training program took place and for the following year. However, when exploring heterogeneous treatment effects, we find that the program increased test scores by 0.11 standard deviations in public schools compared to government-aided schools. This may be driven by two channels: First, public schools are typically located in more rural and economically disadvantaged areas, suggesting that the SLPDP may have been particularly effective in institutions with fewer available resources and less experienced teachers. Students in public schools tend to score approximately 0.03 to 0.05 standard deviations below their counterparts in government-aided schools.

Second, headteachers in government-aided schools operate in slightly different governance structures to headteachers in public schools. Headteachers in government-aided schools may have less time available to manage schools, as they often need to report back to the owners of their schools, frequently affiliated with churches. This puts an extra burden on their time, but also introduces an additional layer to decision-making. This could suggest that implementing meaningful changes within these schools may be inherently more challenging.

While the School Leadership Professional Development Program resulted in only modest improvements in test scores, enhancements of this magnitude can still be cost-effective when compared to alternative educational interventions. One reason for this is that SLPDPs often train only a few actors, in our case only one, per school, rather than a larger number of teachers or students. This aspect is particularly crucial in a resource-constrained country context (Anand et al., 2023; Kremer et al., 2013). Assessing the costs of the training program per trained headteacher and assuming that all students in the school experienced an increase in test scores shows that this SLPDP demonstrates medium cost-effectiveness. However, we argue that if targeted to the right schools or headteachers, this program can become even more cost-effective. As this SLPDP is now being implemented at the national level, further research on the optimal design of SLPDPs, as well as the underlying mechanisms, such as the effects of training on school leadership practices and the effect of these practices on school environments and conditions for teaching and learning, may explain how professional development programs can become more effective. It is equally important to investigate which components make the program in particular efficient and how to reduce overall costs, such as shortening the teaching period or partly teaching online.

This paper contributes to the scarce literature on the effects of better leadership practices on student achievement. To date, the existing literature has shown mixed results when it comes to student achievement (Leaders, 2020; Leithwood et al., 2020). A recent meta-analysis of 14 studies found that school leadership and management programs had a small but positive impact of approximately 0.03 standard deviations on test scores in low- and middle-income countries (Anand et al., 2023). For example, de Hoyos et al. (2020) shows that providing information on student ability to school leaders in Argentina increased test scores in math and reading by 0.35 standard deviations, but also led to changes in curriculum, improved teacher quality assessment, and increased parental awareness. Furthermore, a three-year educational management training program for school principals in Brazil increased student learning by 30% at minimal public cost (Barros et al., 2019).

However, most of the studies did not record any statistically significant learning gains. A randomized school management intervention in Madagascar shows a statistically significant increase in student attendance, but no significant effects on test scores (Lassibille, 2016). In Madhya Pradesh, India, Muralidharan and Singh (2020) evaluated a school leadership intervention that had no effect on management or teaching practices, due to lack of accountability and weak incentives for improvement. Similar interventions in Sri Lanka and Gambia also had no significant impact on student test scores (Aturupane et al., 2022; Blimpo et al., 2015).

Some of these zero treatment effects may be explained by the fact that programs had a low participation rate, frequently falling below 30%, as well as limitations in the number of students taking school exams to detect even small treatment effects with regard to learning outcomes. For example, de Hoyos et al. (2020, 2021) emphasize the consistently low uptake of school leadership training workshops in Argentina. Romero et al. (2022) demonstrate that the already low participation in their SLPDP intervention in Mexico decreased to almost zero, when implemented through a "training of the trainers" method, where key personnel in the school system were trained to subsequently instruct other school actors.

This study addresses these challenges by achieving, compared to other papers, a notable high take-up rate among headteachers (around 70%), as well as a substantial number of test-takers (over 50 000). This allows us to measure precise treatment effects and explore more detailed heterogeneous treatment effects with respect to school location, school type, or headteacher characteristics, creating a deeper understanding of the impact of SLPDP on student achievement. In this regard, the findings of this paper are in line with the recent meta-analysis of Anand et al. (2023) showing small but significant effects on student test scores, but only in public schools in Rwanda. In addition to the empirical evaluation, we also use extensive qualitative interviews with principals and teachers to understand their perceptions of the effectiveness of the program.

In addition, the existing literature on the impact of school leadership professional development programs on educational outcomes in Sub-Saharan Africa is very limited. To our knowledge, only two studies, in Madagascar (Lassibille, 2016) and Gambia (Blimpo et al., 2015), have examined the effects of SLPDPs. We fill this gap by presenting for the first time evidence from a school leadership professional development program in East Africa, Rwanda.

Last, this school leadership professional training program was implemented in districts with a low performance in mathematics and high student dropout rates. In these districts, acquiring information on improving learning outcomes and understanding associated costs is especially crucial with regard to the effectiveness of aid-funded technical assistance programs. We conducted a thorough cost-effectiveness analysis by comparing our detailed data on the program's costs per participant with similar interventions. This helps to determine which interventions are the most cost effective in terms of learning gains for students. It should be noted that so far only four studies from any country have reported exact program costs in school leadership professional development programs (Anand et al., 2023).

2 Intervention

2.1 Educational system in Rwanda

Rwanda basic education consists of six years of primary school, three years of junior secondary education and three years of senior secondary education. Education is compulsory for nine years, from ages 7 to 15 and English is the national language of instruction for all public and government-aided schools. At the end of grade six of primary school, all students sit for the Primary Leaving Examinations (PLE) determining if students can transition to lower secondary education. Students are tested in Mathematics, Science, Social studies, Kinyarwanda and English.

Schools offer varying levels of education, including primary education only, primary and lower secondary education (9YBE) or primary, lower, and upper secondary education (12YBE). Additionally, there are three distinct types of ownership and management structures of schools: public, private, and government-aided schools.[1]

[1] Data from the 2018 Annual School Census reveals that out of 3993 primary schools, 1904 (48 %) were government-aided, 760 (19%) private and 1329 public (33%).

Public schools are both owned and operated by the government, while private schools are typically owned and managed exclusively by non-governmental organizations (NGOs) or private individuals. Government-aided schools are often owned by faith-based organisation, predominately catholic or protestant, where school management involves collaboration between the government and the faith-based organisation (King, ed, 2013; Scheunpflug et al., 2021).

The key distinction between public and government-aided schools lies in the funding structure and governance. Public schools rely predominantly on the government budget, including a capitation grant, regulations and teacher provision. Government-aided schools also receive the capitation grant from the government and teachers are typically on the government payroll. However, they still maintain some autonomy over their operations, such as over the day-to-day operations of the school or teacher appointments (UNESCO, 2023). Overall, government-aided schools often have better infrastructure and tend to perform better academically.

Overall, Rwanda has made commendable strides in education over the past years. Following the 2008 and 2012 policies on nine and twelve years of free basic education, enrollment rates have surged. For instance, the primary gross enrolment ratio (GER) has increased to over 100 percent since 2013. By 2017, practically all primary and secondary schools had toilets and 60 percent had tap water. Hydroelectric supply is available in over 55 percent of primary schools (Trines, 2019).

However, the country's access and the quality of education remain a challenge. While public and government-aided primary and secondary schools are free from tuition, parents still need to pay for mock exam fees, registration fees, etc., and these fees pose serious challenges for successful school attendance, performance, and completion (Trines, 2019). Furthermore, as of 2019, the dropout rate stood at 7.8 percent against a target of 4.3 percent and only 81.6 percent of PLE takers succeed. Results of nationally representative sample-based early grade reading assessments have been consistently poor (Crawfurd, 2021).

2.2 Content of the school leadership professional development program

In 2018, the Ministry of Education in Rwanda adopted the Education Sector Strategic Plan 2018-2024 (ESSP) to integrate English as the primary medium of instruction, implement a competency-based curriculum (CBC) and introduced Information and Communications Technology (ICT) in classrooms

(MOE, 2018).[2] However, a primary concern for the Ministry of Education is the insufficient proficiency of teachers in subject content, pedagogy and English. This could hinder the effective curriculum delivery and may also not positively affect student's learning.

In response to this concern, the Ministry of Education recognized the necessity of transforming the role of school leaders. The goal is to enhance school management, to improve the quality of teaching and learning in schools and to ultimately enhance student learning outcomes. In particular in Rwanda, headteachers are usually promoted from the role of a teacher to a leadership position without receiving any formal training.[3] They may, therefore, feel unprepared and lack the necessary skills to effectively lead the school. This observation is in line with recent observation that most school leaders receive either no or less than two days of leadership training support per year in low-or middle-income countries (Lopez and Rugano, 2018; UNESCO, 2020).

In pursuit of these objectives, the Ministry of Education joined forces with development partners, particularly "VVOB - education for development", to implement a Continuous Professional Development (CPD) program aiming at improving competences of school leaders. The intervention is an accredited diploma program focusing on "Effective School Leadership". The diploma program is delivered and accredited by the University of Rwanda – College of Education for primary and Secondary Schools. The objective is to equip school leaders with essential skills and knowledge required to lead their schools effectively, improve teaching and ultimately improve student outcomes.

The educational curriculum is designed around the five professional standards for effective school leadership, encompassing vital aspects such as (1) creating strategic directions for the school, (2) leading learning, (3) leading teaching, (4) managing the school as an organization and (5) working with parents and the wider community (REB, 2020; Saux et al., 2021).

Overall, the diploma program comprises 40 credits, divided into four modules (10 credits per module). The first module gives an overview of school leadership and focuses on the standard on working with parents and the role of the wider community. In this line, school leaders are instructed on

[2] The Ministry of Education, along with its implementing bodies, the Rwanda Basic Education Board (REB) and the Higher Education Council (HEC) is responsible for developing strategies and national programs and coordinating collaboration with international partners. Between 2011 and 2020, the REB had been responsible for nationwide examinations, the responsibility was then transferred to the National Examination and School Inspection Authority (NESA).

[3] To the best of our knowledge, formal and long term leadership courses were only available through four master's programs in Rwanda: the "Master of Education" at the University of Rwanda College, the "Master in Education Management and Administration" at the University of Kigali, the "Master in Educational Planning and Management" at Mount Kigali University and the "Master of Education degree in Educational Administration" at the Adventist University of Central Africa. However, less than two percent of the trained headteachers in our intervention hold a master's degree, indicating a lack of prior training in school management before our intervention.

how they can apply effective communication and collaboration skills to engage with the community. Building and maintaining relationships with students, teachers, parents and the community is indeed essential for achieving both school and system goals.

The second module teaches how to create strategic direction for the school. It provides headteachers with practical guidance on how to work in collaboration with the school community and diverse stakeholders, such as students, staff, parents, local leaders and development partners, to formulate a common vision, mission, values and strategies for improving the school environment.

The third module discusses how to manage the school as an organization. More precisely, it focuses on resource optimization and how to create a safe and efficient environment for teaching and learning.

The fourth module delves into strategies for both leading learning and teaching. In leading learning, school leaders receive practical insights into establishing a secure and inclusive environment for students, fostering optimal learning conditions. In guiding teaching, leaders are equipped with directives on supporting educators through ongoing feedback and personalized professional development. This ensures that teaching maintains high standards of rigor, relevance and evidence-based practices that align well with the competency-based curriculum.

In essence, the diploma program was designed to initiate change by equipping school leaders with the skills necessary for a distributed leadership model, fostering shared decision-making and collaborative problem-solving within the entire school community.

We anticipate that this further leads to improved teaching methodologies, enriched learning environments and ultimately superior student outcomes. Consequently, this paper aims to investigate whether students enrolled in schools that actively engage in the Continuous Professional Development (CPD) program demonstrate improved performance in national exams and under what specific circumstances this improvement is most notable.

2.3 Roll-out of the school leadership professional development program

The diploma program for school leadership was piloted in 416 primary and secondary schools in 2015/16 (referred to as cohort 0) with the aim of training at least one headteacher per sector[4] in both

[4] Rwanda is divided into four provinces and the city of Kigali, which are further divided into 30 districts. All districts are composed of 416 sectors consisting of 2148 cells and 14837 villages.

primary and secondary schools in all 30 districts in Rwanda.[5] The main selection criteria for this particular cohort were the headteacher's proficiency in English and that the school was public or government-aided. District officials were involved in the selection of participants during the pilot phase.

After the initial pilot, the program, hereby referred to as the Continuous Professional Development (CPD) program, was expanded to include three additional cohorts at the primary level.[6] In contrast to the pilot project, the CPD was now implemented in a total of six districts: Kayonza, Kirehe, Gatsibo, Nyabihu, Nyagatare and Rusizi. The districts were chosen based on their dropout rates and math exam scores, aligning with VVOB's objective to improve STEM and math performance and to reduce drop-out rates. A detailed explanation of the selection process can be found in figure A1.

The program aimed to provide training to headteachers in a total of 525 schools, spread across three cohorts.[7] The allocation of schools across the three cohorts was randomly, which enables us to estimate the causal impact of the SLPDP. The Rwanda Basic Education Board supplied VVOB with a list of all 591 primary schools in the six targeted districts in 2017, including information on their ownership (private, government-aided, public) and schooltype (primary, 9YBE, and 12 YBE). Out of the 591 primary schools, 66 had already participated in cohort 0, leaving 525 schools eligible for training. Each of these schools was then assigned a random number between 0 and 1. Subsequently, they were ranked according to that number from the highest to the lowest number. Then, VVOB proceeded by distributing them across the three cohorts in descending order, stratified by district and school-type with the aim of having an equal number of schools per district and school-type in each cohort.

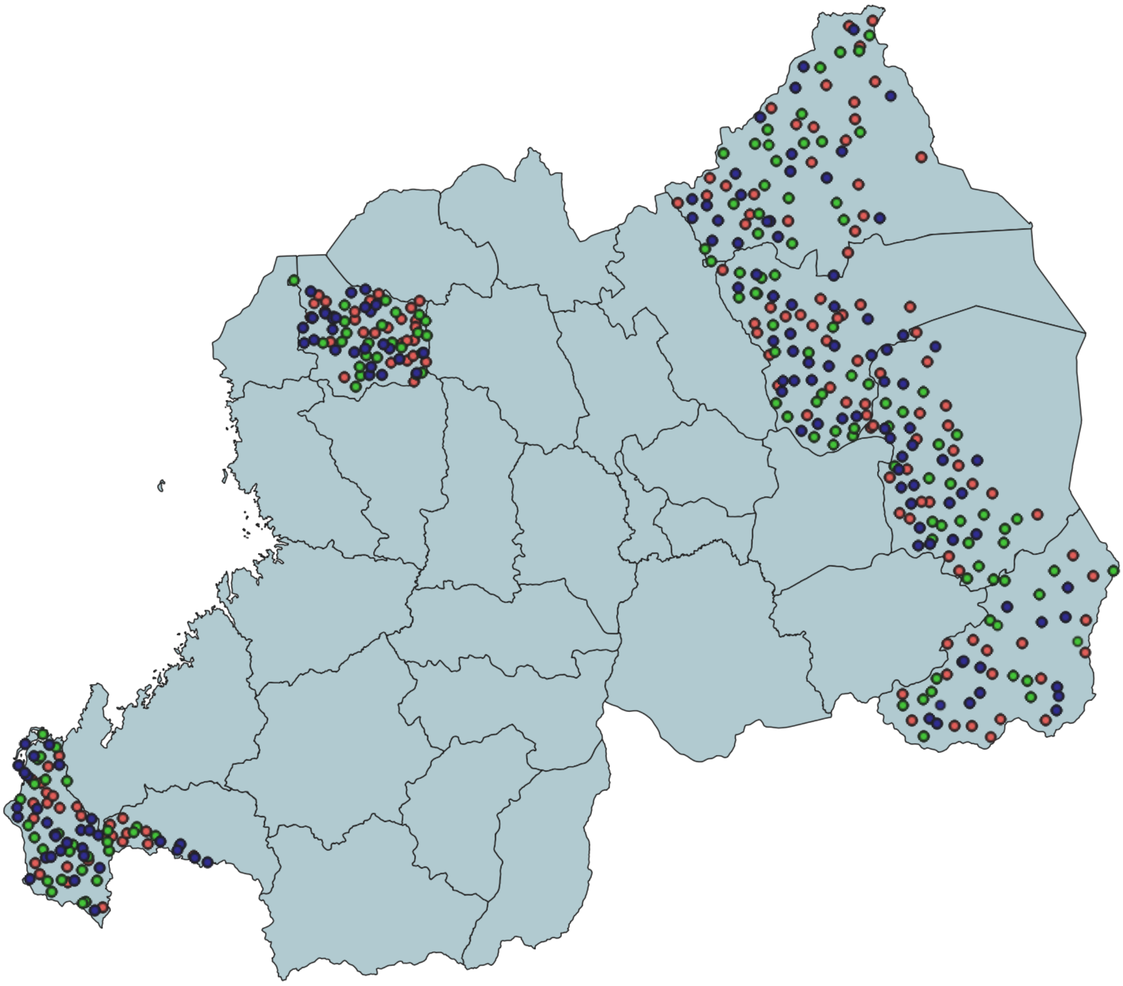
Figure 1 displays the locations of the schools in cohort 1, 2, and 3 in the six chosen districts across Rwanda. In addition, table A1 shows that the schools are evenly distributed across the six districts and with regard to schooltype or ownership.

[5] If a school provides only primary education, the headteacher typically oversees the entire school. However, in schools that offer both primary and secondary education, it is common to have separate headteachers for the primary and secondary sections.

[6] The CPD was funded by the Belgian Directorate-General for Development Cooperation and Humanitarian Aid (DGD) and was implemented by VVOB and partners.

[7] Note that in 2018, with support of the Mastercard Foundation and as part of the 'Leaders in Teaching' initiative, the professional development program has been implemented in 14 districts, with an overlap in 3 districts (Kayonza, Nyabihu and Rusizi) for secondary headteachers. Starting from 2022, the program has been implemented at the national level for primary and secondary headteachers, again with funding of DG.

Figure 1: **Locations of primary schools in cohort 1 (red), cohort 2 (blue) and cohort 3 (green).**



After having compiled a list of schools to be contacted for each cohort, headteachers were then invited to participate in the SLDPD. This implies that headteachers had the option to decide whether to participate or not in the training program. However, the SLPDP's status as an accredited training program, certified by the College of Education - University of Rwanda, served as a strong incentive for participation. Additionally, the training was offered free of charge to the headteachers, with travel costs covered by VVOB.

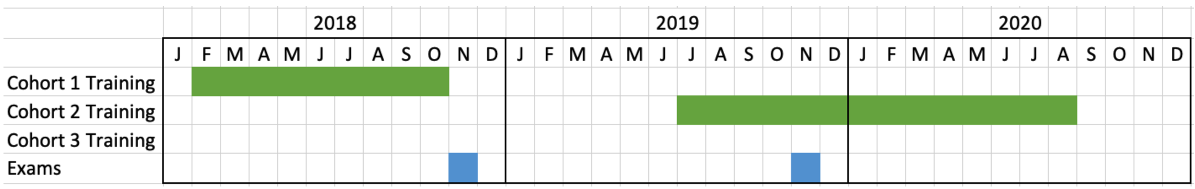
If the initially selected participant is unavailable, the invitation is extended to the headteacher of another school within the same sector, and if needed, within the same district. For the next cohort, the initially selected participant is re-contacted to assess their current availability.

Overall, there is imperfect compliance with the program, as some headteachers selected for cohort 1 participated in cohort 2, some headteachers from cohort 2 participated in cohort 1 and some headteachers from cohort 3 participated in cohort 2 or cohort 1. The uptake, measured by attendance until the completion of the training program varied across the three cohorts as following:

- **2018 (cohort 1):** Of the original 175 schools in cohort 1, 125 schools received the training in 2018, while 9 schools from cohort 2 and 14 schools from cohort 3 received also the training in 2018.
- **2019 (cohort 2):** Out of the initial 175 schools in cohort 2, only 46 schools received the full training in 2019. Furthermore, 37 schools from cohort 3 and 9 schools from cohort 1 underwent training in 2019. The onset of the Covid-19 pandemic significantly contributed to a high dropout rate, given that schools were forced to close and certain aspects of the training were transitioned to an online format. Notably, some schools faced challenges accessing the online training due to a lack of computers.
- **2020 (cohort 3):** The training for cohort 3 in 2020 was temporarily suspended due to the testing of an online training program designed for secondary headteachers. Consequently, the original form of the training was not delivered to any headteacher.

As compliance with the program was imperfect, this paper measures the intent-to-treat effect. The timeline illustrated in figure 2 summarizes the key events. The training for cohort 1 ran from February to October 2018 and for cohort 2 from July 2019 to August 2020.[8] The third cohort was scheduled to receive training in 2020, but due to the Covid-19 pandemic, none of the primary schools in this cohort were able to participate in the training as planned in 2020, but received the training in a blended modality in 2021.

Figure 2: Timeline of the program



For the first cohort the program was setup to have 18 contact days, including two examination days. The remaining 16 days were taught in blocks of two days. In 2019 the program was offered as a blended program, with 14 training days face-to-face and 2 days through online learning. In general, headteachers received the training in-person at local training centers. The instructors would then

[8]Because of the Covid-pandemic and the online transition of some training modules, the total length of the training for cohort 2 increased.

move on to another district and give the training there. Overall, there is minimal turnover among instructors ensuring that all headteachers experienced a similar level of training quality.

We examine whether the School Leadership Professional Development Program (SLPDP) had an impact on test-scores of the Primary Leaving Examinations (PLE). Figure 2 illustrates that cohort 1 completed the training before the 2018 exams, while cohort 2 received partial training before the 2019 exams. To obtain a more accurate estimate, we exclude cohort 2 from most of our analyses and only compare the effect of SLPDP on test-scores between cohort 1 and cohort 3.

This implies that in our main specification, the treatment group are the 175 schools from cohort 1 and the control group are the 175 schools from cohort 3. The sample size is sufficient to detect a treatment effect of at least 0.05 standard deviations with respect to test-scores.[9]

3 Identification

3.1 Data

We use various data sources for this paper. First, information on the selected schools (in cohorts 1, 2, and 3), and information on participating school leaders were provided by the VVOB. Data on the program participants include birth year, gender, education level, training dates and if they passed the final exam.

The second dataset used in the study consists of student test score records from the P6 primary school national exam conducted between 2015 and 2019. These test scores are administered and recorded by the National Examination and School Inspection Authority (NESA). In addition to students' test scores, the dataset includes school-names, geolocations of the schools, as well as students' year of birth and gender.

Third, we rely on other complementary data-sets, such as the annual school census data from 2017 and 2018. We leverage these datasets to gather additional information on school characteristics prior to the program implementation. We obtained information on whether the school has access to electricity and water and on the number of teachers, students, students in P6 and classrooms per school. We also rely on Prio Grid data to enhance school characteristics with information on nightlight activity per sector, gross cell product per sector and the distance to Kigali. Furthermore, we use the

[9] A post-treatment power calculation reveals that 54 treatment and 54 control schools are required to detect an effect that is larger than 0.05 standard deviations for test-scores, given that we have a type I error rate of 0.05, desired power of 0.80, inter-cluster correlation of 0.05, 100 students on average per school in P6, 3 strata and an average test-score variance of 0.80 standard deviations computed with test-score data from the Rwanda PLE exams in 2017.

population census of 2012 to include the population density of the sector where the school is located.[10]

The process of merging the annual school census data with the program information provided by VVOB was carried out using official school and district codes. Having compiled a dataset with program information and school census information, we then merged it with the offical PLE test-score data. This was challenging, as the PLE data-set contained the school names, but not valid school codes. Therefore, we used fuzzy matching based on the school name, sector name and district name. Only pairs with at least an 80% match were kept. Following this, the data was manually cleaned for further accuracy.

Out of the 274 government-aided schools that were randomly assigned to the treatment or control group, we successfully matched 255 schools (93%) with the test-score data from the National Examination and School Inspection Authority (NESA). Similarly, out of the 171 public schools that were randomly assigned, we matched 156 schools (91%) with the test-score data. However, out of the 80 private schools that were randomly assigned, we only matched 7 schools (9%) with the test-score data. One possible explanation of such a low matching rate for private schools is that many private schools may not necessarily follow the national curriculum and do not have had a grade six at the time of the intervention. Therefore, we decided to exclude the private schools from the analysis. This implies that our study focuses solely on the public and government-aided schools in Rwanda. We will, therefore, conduct our balancing test on public and government-aided schools solely.[11]

The final data-set appears as following: Out of the 149 public and government aided schools from cohort 1 schools, 137 were successfully merged. Similarly, in cohort 2, 132 out of the total 145 public and government aided schools were merged and in cohort 3, 138 out of the total 149 public and government aided schools were merged. Table 1 gives an overview how many public and government aided schools were merged with the test-score datasets and how many of them actually participated and completed the training program in 2018, 2019 and 2020.

[10] Population density measures the average number of inhabitants per square kilometer within a specific area. The data on population density is sourced from the fourth Population and Housing Census conducted in Rwanda in 2012. It is important to note that the latest population density information was gathered in 2022, so after the intervention took place.

[11] However, given that there is roughly the same attrition in each cohort concerning public and government-aided schools, that the program was randomized and that baseline characteristics are controlled for, the missing private schools should not bias the estimation.

Table 1: **Compliance rates by cohort in the final data set for public and government-aided schools**

	Cohort 1	**Cohort 2**	**Cohort 3**
Merged	136/149 (90%)	132/145 (92%)	140/149 (95%)
Compliance with the program:			
Received training in 2018	99/136 (73%)	6/132 (5%)	10/140 (7%)
Received training in 2019	7/136 (5%)	40/132 (30%)	33/140 (23%)
Received training in 2020	0/136 (0%)	0/132 (0%)	0/140(0%)

3.2 Estimation strategy

As outlined, this paper measures the average intent-to-treat effect of the program on student's PLE test-scores. The main estimation is as following:

$$Y^j_{isd} = \alpha + \beta \times Treatment_s + X_{isd} + \bar{Y}^j_{s,t-1} + Z_s + \lambda_d + \epsilon_{isd}, \tag{1}$$

where the outcome of interest, denoted as Y^j_{isd}, represents the individual student's performance i in subject j on the primary leaving examination in 2018 or 2019 at school s in district d. Given the program's focus on improving STEM grades, we compare STEM to non-STEM grades. Considering that the Ministry of Education instructed English to be the primary language of instruction only one year before the SLPDP was implemented, we also examine English grades. In addition, to assess the overall effects, we analyze aggregated test scores.[12] We standardize the grades in the treatment and control group with mean 0 and a standard deviation of 1.[13]

$Treatment_s$ is a dummy and equals one if a school was assigned to the first cohort otherwise 0. This implies that the treatment group consists of schools that were assigned to cohort 1 and the control group are schools that were assigned to cohort 3. Student covariates X_{isd} are gender and birth year. School covariates Z_s consist of a dummy indicating if a school is a government-aided school and a categorical variable to control for school-type (Primary, 9 YBE, 12YBE). To take into account prior school performance, we further include the average grade in the subject of interest per school one year before the intervention took place, $\bar{Y}^j_{s,t=-1}$.[14]

[12]Aggregated grades consist of Mathematics, English, Sciences, Social Sciences and Kinyarwanda grades. Stem grades are Sciences and Mathematics grades, while non-stem grades are English and Social Sciences grades.

[13]Exams in Rwanda were graded between 1 and 9 in Rwanda until 2019, where 1 is the best and 9 the worst grade. We standardize those grades and multiply the standardized values by -1 implying that a one standard deviation increase refers to a better test-score.

[14]We do not have the grade information for 4 schools in 2017 and 2016. We replace this missing information with the average grade in subject *j* from 2017 or 2016.

We employ the lasso approach for prediction and model selection to enhance the precision of our estimates by incorporating additional control variables. Through lasso, we identify the following control variables as highly relevant: the population density of the school's sector, whether a school was opened after the Tutsi genocide in 1994, the availability of electricity, the total number of students and teachers, the number of students taking the PLE exam in 2017, the distance to the capital and nightlight activity within the school grid. Nightlight activity serves as a proxy for economic activity within the school grid and can be considered a measurement of wealth (Bruederle and Hodler, 2018; McCord and Rodriguez-Heredia, 2022).[15]

For the heterogeneity analysis, we add an interaction term between $Treatment_s$ and a variable of interest. More precisely, we interact $Treatment_s$ with the gender of the student, the school-ownership and whether the sector the school is located in a sector below or above the median nightlight-activity and population density.[16] In the heterogeneity analysis section, we explain why we selected these variables of interest.

We address the issue of multiple hypothesis testing by adjusting the p-values using the false discovery rate method introduced by Benjamini and Hochberg (1995). The p-values are corrected for each treatment arm and outcomes are grouped into families according to the specifications outlined in the table notes.

In light of two-sided non-compliance with the program, as some schools from cohort 1 participated in the training program for cohort 2 and some schools from cohort 3 participated in the training program in cohort 2 or cohort 1, we also estimate the local average treatment effects (LATE). LATE is the effect of the SLPDP on headteachers that complied with the treatment. In order to run LATE, we as-

[15]We have information on the number of teachers and students for 90% of schools. Additionally, data on the number of students in P6 and classrooms in P6 is available for 72% of schools. Rather than dropping schools with missing information, we impute the missing data using sample averages and include a dummy variable in the regression to indicate if a school has missing information.9

[16]Please note that the classification of schools into 'urban' or 'rural' categories is based on a comparison of the population density within the school's sector against the median population density per sector. A school's sector is deemed 'urban' if its population density exceeds the median. It is important to also note that while our test-score data includes a variable indicating whether a school is located in an urban or rural area, the representation of urban schools within our cohorts is too minimal. Specifically, only 9 schools from cohort 1 and 8 schools from cohort 3 are classified as urban.

sume that the non-interference[17], the excludability[18] and monotonicity[19] assumptions are satisfied. In a first step, the random allocation across cohorts is used as an instrument for actual participation in the school leadership training program:

$$P_s = \alpha + \beta \times Treatment_s + X_{isd} + \bar{Y}^{j}_{s,t-1} + Z_s + \lambda_d + \epsilon_{isd}, \tag{2}$$

where P_s equals one if the headteacher participated in the program and zero otherwise. Subsequently, the predicted implementation is used in a second step to estimate the effects on student's test-scores:

$$Y^{j}_{isd} = \alpha + \beta \times \tilde{P}_s + X_{isd} + \bar{Y}^{j}_{s,t-1} + Z_s + \lambda_d + \epsilon_{isd}, \tag{3}$$

4 Results

4.1 Balancing and summary statistics

This section presents summary statistics and assesses the balance between schools in cohort 1 (treatment group) and cohort 3 (control group) within our final school sample, so public and government-aided schools that could be successfully matched to the official test-score data. VVOB did not collect independent baseline data before the randomization was conducted, being the reason why we rely instead on detailed administrative school census and test-score data.[20]

Schools in cohort 1 (treatment group) exhibit a well-balanced distribution of school and student characteristics, when compared to schools in cohort 3 (control group). Looking at school characteristics in table 2 reveals that there is no imbalance with respect to school-type, school-ownership or

[17] This assumption posits that the treatment assignment of one headteacher (denoted as i) does not influence the treatment status of another headteacher (denoted as j). In this intervention, headteachers were organized into cohorts and participation invitations were extended accordingly. Therefore the treatment status of i does not impact the treatment status of j. This non-interference principle ensures independence in treatment assignment decisions.

[18] The excludability assumption asserts that the treatment assignment and treatment status of other headteachers j do not affect the outcomes for a specific headteacher i after participating in the SLPDP. This assumption holds true as we rule out the possibility of spill-over effects.

[19] The monotonicity assumption states that if a headteacher is moved from one cohort to another (e.g., from cohort 1 to cohort 3), their treatment status should either remain unchanged or compliance with the treatment should increase. While a formal test for this assumption may not be feasible, we posit that headteachers who did not participate in the SLPDP were likely not motivated or constrained by time. This implies that a headteacher who is not motivated will neither participate in cohort 1 nor in cohort 2. At the same time, a headteacher who is motivated, but does not have time to participate in cohort 1, will participate in cohort 2. Thus, moving headteachers from cohort 1 to cohort 2 or vice versa would either leave the treatment status unchanged or compliance with treatment may increase.

[20] In fact, the existence of baseline administrative data for annual school census data and test-scores allows us to conduct the experimental evaluation without an independent baseline, as outlined in (Muralidharan, 2017). This approach also ensured a prudent use of research funds given the risk of either non-implementation or non-compliance with the RCT protocol by VVOB and the Ministry of Education.

the number of classrooms in P6, among other variables tested.

However, schools in the control group are situated in slightly more densely populated areas and more students appear to take the PLE exam in the control group. For instance, in 2017, the mean difference between students taking the PLE exam between control and treatment group was 15, significant at the five percent level. In 2016, the difference was eight students, but not significant though. Therefore, we include population density, as well as the number of PLE test-scores, when running the main regression with additional controls. Overall, the F-test of joint significance is insignificant with an F-value of 0.93 and a p-value of 0.56.

We further test if there is a significant difference in test-scores, gender and birth year of students in treated and control schools that took the PLE exam in 2017, 2016 and 2015. Table 3 shows that there is no imbalance between control and treatment schools with regard to the birth year, the gender and subject grades. In addition, the F-test of joint significant is insignificant for test-scores in 2017, 2016 and 2015.

Table 2: **Baseline descriptive and balance**

Variable	Sample mean	Treatment	Control	Mean diff.	P-value	Size treatment	Size control
School characteristics							
Government aided 2018	0.61	0.60	0.62	-0.02	0.67	136	140
	(0.48)	(0.49)	(0.49)				
Primary 2018	0.61	0.60	0.62	-0.02	0.77	136	140
	(0.49)	(0.49)	(0.48)				
9 YBE 2018	0.23	0.25	0.21	0.03	0.45	136	140
	(0.42)	(0.43)	(0.41)				
12 YBE 2018	0.16	0.15	0.16	-0.02	0.85	136	140
	(0.36)	(0.37)	(0.37)				
Total students in 2017	1233.14	1178.69	1286.04	-115	0.18	136	140
	(769.12)	(706.40)	(824.60)				
Students in P6 in 2017	112.33	108.38	116.16	-8.29	0.28	136	140
	(68.59)	(67.18)	(69.96)				
PLE takers in 2017	89.29	71.74	96.63	-14.78	0.02**	136	140
	(57.40)	(49.30)	63.62				
PLE takers in 2016	73.88	69.63	78.02	-8.04	0.15	136	140
	(48.22)	(43.21)	(52.46)				
P6 classrooms in 2017	2.67	2.76	2.59	0.17	0.51	136	140
	(2.20)	(2.63)	(1.69)				
Teachers in 2017	19.00	18.15	19.84	-1.80	0.22	136	140
	(13.08)	(12.54)	(13.58)				
Establishment date	1980	1982	1980	1.41	0.63	136	140
	(24.56)	(24.64)	(24.56)				
Establishment after 1994	0.43	0.46	0.41	0.05	0.35	136	140
	(0.49)	(0.41)	(0.46)				
Electricity 2017	0.50	0.48	0.52	-0.03	0.56	136	140
	(0.50)	(0.50)	(0.50)				
Water 2017	0.55	0.54	0.56	-0.02	0.71	136	140
	(0.50)	(0.50)	(0.50)				
School feeding program	0.30	0.32	0.28	0.04	0.37	136	140
	(0.46)	(0.47)	(0.45)				
Distance to the capital in km	233	228	239	-0.61	0.99	136	140
	(464.61)	(454.76)	(475.70)				
Gross Cell product in 2005	0.12	0.12	0.12	0.00	0.93	136	140
	(0.06)	(0.06)	(0.06)				
Night light activity in 2012	0.96	0.89	1.03	-0.03	0.77	136	140
	(1.34)	(1.35)	(1.34)				
Night light activity in 2012 above median	0.51	0.52	0.50	0.01	0.85	136	140
	(0.50)	(0.50)	(0.50)				
Urban	0.07	0.07	0.07	0.00	0.81	136	140
	(0.26)	(0.26)	(0.26)				
Pop density 2012	463.40	429.43	496.40	-59.10	0.03**	136	140
	(463.40)	(429.43)	(496.40)				
Pop density above median	0.50	0.45	0.55	-0.08	0.11	136	140
	(0.50)	(0.50)	(0.50)				
F-test of joint significance	F-value: 0.81	P-value: 0.70					

NOTE: Significance levels: *<0.10; **<0.05 ***<0.01. The values in brackets denote standard deviations. The mean difference is computed by regressing treatment on the covariate including district fixed effects. Standard errors are clustered at the school level. All grades are standardized with mean 0 and standard deviation 1. There are 140 control and 136 treatment schools for student characteristics in 2017. A more detailed description of the variables can be found in table A2.

Table 3: **Baseline descriptive and balance (continued).**

Variable	Sample mean	Treatment	Control	Mean diff.	P-value	Size treatment	Size control
Student characteristics 2017							
Aggregate grade 2017	0.00	0.06	-0.05	0.09	0.13	11117	13440
	(1.00)	(1.05)	(0.99)				
Stem grade 2017	0.00	0.04	-0.03	0.06	0.28	11117	13440
	(1.00)	(1.00)	(0.99)				
Non-Stem grade 2017	0.00	0.05	-0.04	0.08	0.22	11117	13440
	(1.00)	(1.01)	(0.98)				
English grade 2017	0.00	0.04	-0.03	0.06	0.29	11117	13440
	(1.00)	(1.01)	(0.98)				
Birth year	2002	2002	2002	-0.00	0.67	11117	13440
	(1.68)	(1.72)	(1.66)				
Female student	0.54	0.54	0.55	-0.01	0.36	11117	13440
	(0.49)	(0.49)	(0.49)				
F-test of joint significance	F-value: 1.27	P-value: 0.27					
Student characteristics 2016							
Aggregate grade 2016	0.00	0.04	-0.04	0.01	0.84	9397	10923
	(1.00)	(0.98)	(1.01)				
Stem grade 2016	0.00	0.00	-0.00	-0.00	0.95	9397	10923
	(1.00)	(0.97)	(1.02)				
Non-Stem grade 2016	0.00	0.01	-0.01	0.01	0.84	9397	10923
	(1.00)	(0.99)	(1.00)				
English grade 2016	0.00	0.01	-0.01	0.02	0.78	9397	10923
	(1.00)	(0.98)	(1.01)				
Birth year	2001.57	2001.54	2001.58	-0.02	0.65	9397	10923
	(1.64)	(1.62)	(1.65)				
Female student	0.55	0.54	0.55	-0.00	0.63	9397	10923
	(0.49)	(0.50)	(0.49)				
F-test of joint significance	F-value: 0.34	P-value: 0.92					
Student characteristics 2015							
Aggregate grade 2015	0.00	0.03	-0.03	0.05	0.31	8183	9434
	(1.00)	(0.98)	(1.01)				
Stem grade 2015	0.00	0.01	-0.01	0.02	0.65	8183	9434
	(1.00)	(0.97)	(1.02)				
Non-Stem grade 2015	0.00	0.04	-0.03	0.07	0.24	8183	9434
	(1.00)	(0.99)	(1.00)				
English grade 2015	0.00	0.02	-0.02	0.04	0.50	8183	9434
	(1.00)	(0.98)	(1.01)				
Birth year	2001.58	2000.58	2001.58	0.00	0.98	8183	9434
	(1.64)	(1.62)	(1.65)				
Female student	0.55	0.54	0.54	0.00	0.74	8183	9434
	(0.49)	(0.50)	(0.49)				
F-test of joint significance	F-value: 0.94	P-value: 0.47					

NOTE: Significance levels: *<0.10; **<0.05 ***<0.01. The data from this table is obtained from the test-score dataset. The values in brackets denote standard deviations. The mean difference is computed by regressing treatment on the covariate and district fixed effects. Standard errors are clustered at the school level. All grades are standardized with mean 0 and standard deviation 1. Overall, there are 136 schools in the treatment group and 140 schools in the control group.

4.2 Average treatment effects

This section reports the average intent-to-treat effects, before discussing heterogeneity. It is important to note that the SLPDP did not significantly the composition of students attending the treatment schools and therefore the presented estimates are not likely to be biased downwards. We address this issue in the placebo section of this paper.

Table 4 presents the effects of the school leadership professional development program on test-scores in 2018. As previously indicated, all coefficients are measured in terms of standard deviations and show if test-scores for cohort 1 increased or decreased following the SLPDP. The treatment effects are mostly positive, but small in size and insignificant across all columns. As the coefficients in columns (1) to (8) are estimated with relatively small standard errors, we conclude that the SLPDP intervention has a precise zero effect for test-scores in 2018.

Table 4: **Treatment effects on test-scores in 2018.**

Variable	(1) Aggregate	(2) Aggregate	(3) English	(4) English	(5) Nonstem	(6) Nonstem	(7) Stem	(8) Stem
Treatment	0.01	-0.01	0.02	0.00	0.03	0.02	0.01	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Add. controls	no	yes	no	yes	no	yes	no	yes
Number of students	26871	26871	26871	26871	26871	26871	26871	26871
Number of schools	276	276	276	276	276	276	276	276
R^2	0.20	0.21	0.17	0.18	0.20	0.21	0.17	0.18

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership, the average grade per school in 2017 and population density of the school's sector. Additional controls include, whether a school was opened after 1994, electricity, total number of students and teachers, number of students taking the PLE exam in 2017, distance to the capital and nightlight activity. If information is missing for the additional controls, we replace the missing value with the average across all schools and include a dummy indicating whether a school has missing information. Standard errors are clustered at the school level.

It is, however, possible that the treatment effects only manifest a year after. Such delays could be attributed to various factors, such as the inability of headteachers to implement new structures or to recruit additional teachers until the next academic year. We test this by estimating the effect of the program on test-scores in 2019, as shown in table 5. The point estimates are higher than those in 2018. For instance, the average treatment effect for aggregate test-scores is 0.03 standard deviations in column (1) and (2), while the ATE was 0.01 and even -0.01 in 2018.[21] However, none of the point estimates are statistically significant.

In summary, this section shows no meaningful effect of the intervention on test-scores, either in the short- or longer-term, when estimating average-treatment effect. The results, may however, still

[21] The estimated coefficients in table 5 are comparable to the average effect size of 0.04 standard deviations reported by the meta-analysis of school leader training programs by Anand et al. (2023).

suggest that the SLPDP could have been effective in the longer-run. Also, as some schools of cohort 3 received partial training in 2019, this could potentially introduce a downward bias in the observed treatment effects.

Table 5: **Treatment effects on test-scores in 2019.**

Variable	(1) Aggregate	(2) Aggregate	(3) English	(4) English	(5) Nonstem	(6) Nonstem	(7) Stem	(8) Stem
Treatment	0.03	0.03	0.03	0.03	0.02	0.02	0.04	0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Add. controls	no	yes	no	yes	no	yes	no	yes
Number of students	30438	30438	30438	30438	30438	30438	30438	30438
Number of schools	276	276	276	276	276	276	276	276
R^2	0.18	0.18	0.15	0.15	0.17	0.18	0.16	0.16

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership, the average grade per school in 2017 and population density of the school's sector. Additional controls include, whether a school was opened after 1994, electricity, total number of students and teachers, number of students taking the PLE exam in 2017, distance to the capital and nightlight activity. If information is missing for the additional controls, we replace the missing value with the average across all schools and include a dummy indicating whether a school has missing information. Standard errors are clustered at the school level.

4.3 Heterogeneous effects

We explore heterogeneity in the treatment effects with respect to the student's gender, the school's ownership (public or government-aided) and the sector characteristics of the school in terms of population density and nightlight activity. Below, we briefly describe why we are interested in these particular variables.

We focus on gender dynamics among students, given the contrasting dynamics still at play in Rwanda. While there is gender parity in primary-level enrollment, boys tend to enroll at a later age than girls and experience higher dropout and repetition rates. Consequently, girls progress through the system more quickly overall, but they continue to achieve lower learning outcomes. The exact mechanism are not fully understood yet, but male students may have an advantage over female students, as female students face more school absences due to increased family responsibilities or a lack of sanitary pads during menstruation(MOE, 2018; Nzaramba et al., 2021; UNICEF, 2024). Our focus revolves around investigating the potential impact of the SLPDP in enhancing female test scores relative to male test scores.

Moreover, as outlined in section 2, public or government-aided schools have a different funding and governance structure. Given these differences, our aim is to analyze if there are differential effects between public and government-aided schools resulting from the SLPDP. It may be the case that the effectiveness of such interventions varies depending on the structure in which the school operates.

Also, the impact of school leadership management interventions may vary based on the resources a school possesses and its initial management level (Anand et al., 2023; Muralidharan and Singh, 2020). We posit that schools situated in less densely populated areas or areas with limited nightlight activity may inherently possess fewer resources and comparatively weaker management capabilities. For instance, schools in sparsely populated areas often encounter challenges in attracting and retaining qualified personnel, obtaining adequate funding and maintaining robust infrastructure (MOE, 2018; Scheunpflug et al., 2021). Moreover, limited nightlight activity serves as an indicator of reduced economic activity and community resources, which, in turn, may further diminish the resources available to a school. When interpreting the results, please note that there is a higher concentration of public schools with approximately 65%, in sectors below median nightlight activity and population density.

We perform the heterogeneity analysis on test-scores in 2018 and 2019 to increase power and precision. Table 6 displays the impact of the treatment variable on the base category compared to the interaction term. For instance, in column (1), "Treatment" indicates how the SLPDP affected male students, while "Treat. × Female" depicts how the treatment effect differed between male and female students. Similarly, in column (2), "Treatment" shows how the intervention affected public schools, while "Treat. × Government Aid." indicates the difference in the effect for government-aided compared to public schools. In columns (3) and (4), "Treatment" refers to schools located in sectors with below-median nightlight activity or below-median population density, while the interaction terms show how the effect changes for schools located in sectors with above-median population density and above-median nightlight activity. We conduct these regressions on aggregated test-scores in columns (1)-(4), stem test-scores in columns (5)-(8) and non-stem test-scores in columns (9)-(12).

To begin with table 6 confirms previous research that highlight a significant and persistent gender gap in learning outcomes between female and male students. More specifically, male students outperform female students by 0.20 to 0.24 standard deviations at the one percent significance level. Relating to this, the SLPDP did not succeed in reducing this gap. If there is any effect male test-scores increased by 0.05 standard deviations, while there is no impact on female test-scores, as reported in column (5).

However, columns (2), (6) and (10) show that the SLPDP significantly increased aggregate, stem, and non-stem test scores by 0.11 standard deviations in public schools. There appears to be no treatment effect for government-aided schools. The interaction terms "Treat. × Government aided" are negative and significant suggesting that the overall effect of the SLPDP for government-aided schools is close to zero.

Furthermore, in sectors below the median nightlight activity and population density, aggregate test-scores may also have increased by 0.02 to 0.06 standard deviations, even though the coefficients are not significant. In sectors above the median nightlight activity and population density, the overall treatment effect is close to zero, as shown in columns (3) and (4).

Table 6: **Heterogeneous treatment effects on test-scores in 2018 and 2019.**

Variable	(1) Ag.	(2) Ag.	(3) Ag.	(4) Ag.	(5) Stem	(6) Stem	(7) Stem	(8) Stem	(9) Non	(10) Non	(11) Non	(12) Non
Treatment	0.03	0.11**	0.06	0.02	0.05	0.11*	0.06	0.04	0.03	0.11*	0.06	0.01
	(0.04)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.05)
Female student	-0.21$^{***}_{***}$				-0.25$^{***}_{***}$				-0.21$^{***}_{***}$			
	(0.02)				(0.02)				(0.02)			
Treat. × Female student	-0.02				-0.05*				-0.02			
	(0.02)				(0.03)				(0.02)			
Government aid.		0.04				0.04				0.03		
		(0.05)				(0.05)				(0.05)		
Treat. × Government aid.		-0.16**				-0.14**				-0.14*		
		(0.07)				(0.07)				(0.08)		
Nightlight ≥ Med.			0.05				0.07				0.04	
			(0.07)				(0.05)				(0.07)	
Treat. × Nightlight ≥ Med.			-0.08				-0.08				-0.08	
			(0.07)				(0.06)				(0.07)	
Pop density ≥ Med.				-0.01				-0.00				-0.06
				(0.05)				(0.06)				(0.05)
Treat. × Pop density ≥ Med.				-0.02				-0.03				0.03
				(0.06)				(0.06)				(0.07)
Test year fe	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of students	57309	57309	57309	57309	57309	57309	57309	57309	57309	57309	57309	57309
Number of schools	276	276	276	276	276	276	276	276	276	276	276	276
R^2	0.19	0.19	0.19	0.19	0.16	0.16	0.16	0.16	0.18	0.19	0.19	0.19

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. Reported are aggregated test-scores in columns (1)-(4), stem test-scores in columns (5)-(8) and non-stem test-scores in columns (9)-(12). The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership and the average grade per school in the given subject in 2017. Standard errors are clustered at the school level. Reported are unadjusted p-values** and adjusted p-values$_{*}$ for multiple hypothesis testing. All outcomes variables in the table are grouped as one family.

However, it is crucial to acknowledge that the significance levels observed for public schools diminish when adjustments are made for multiple hypothesis testing. As a result, the estimated treatment effects become only weakly identifiable. Consequently, to support the argument that the SLPDP might have marginally increased test scores in public schools, as well as in less affluent and more remote areas, we provide additional evidence.

In this regard, we apply a quadratic polynomial fit of degree 2 that explores the correlation between population density (divided into 20 quintiles), nightlight activity and test-scores measured in standard deviations. Figures A2 and A3 show the locally weighted scatterplot smoothing and reveal how standardized grades change for both the control and treatment group with an increase in population density or nightlight activity. The non-linear treatment effects show a significant difference in test-scores in areas with less nightlight activity and population density. As population and nightlight activity increases, this gap gradually narrows and becomes statistically insignificant.

As a result, the SLPDP program may have demonstrated greater effectiveness in public schools situated in rural and less affluent areas, in contrast to government-aided schools located in wealthier and more urbanized regions. We present additional placebo tests that support this hypothesis, along with further explanations in the mechanism section of this paper.

4.4 Local average treatment effects

In this section, we estimate the local average treatment effects (LATE) of receiving the treatment in cohort 1. We use an instrumental variable approach, where the instrument is the assignment to the treatment in cohort 1. In theory, these estimates should be higher than the intent-to-treat estimates to argue that the program actually worked. We run LATE separately on test-scores in 2018 and 2019.

Table 7: **Local average treatment effects on test-scores in 2018.**

Variable	Aggregate	Aggregate	English	English	Nonstem	Nonstem	Stem	Stem
Treated	0.02	-0.01	0.03	0.00	0.05	0.02	0.02	-0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
First-stage								
Treatment assignment	0.67***	0.66***	0.67***	0.66***	0.67***	0.66***	0.66***	0.66***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
H_0: weak instrument								
Robust F-statistic	167.77	165.49	173.93	168.19	174.70	168.98	164.94	163.30
Probability > F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Add. controls	No	Yes	No	Yes	No	Yes	No	Yes
ITT 2018	0.01	-0.01	0.02	0.01	0.03	0.02	0.01	-0.00
Number of students	26871	26871	26871	26871	26871	26871	26871	26871
Number of schools	276	276	276	276	276	276	276	276
R^2	0.20	0.21	0.17	0.18	0.20	0.21	0.17	0.18

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership, population density of the sector where the school is located in and the average grade per school in 2017. Additional controls refer to, whether a school was opened after 1994, electricity, total number of students and teachers, number of students taking the PLE exam in 2017, distance to the capital and nightlight activity. If information is missing for the additional controls, we replace the missing value with the average across all schools and include a dummy indicating whether a school has missing information. Standard errors are clustered at the school level. ITT records the intent-to-treat effect as estimated previously for 2018.

Tables 7 and 8 report the results of the LATE estimates. The robust F-statistic in both tables is well above the common threshold of ten, suggesting that our instrument is significant. Moreover, the LATE effects are either the same or about 0.01 to 0.02 standard deviations higher than the ITT effects across all specifications. Although the estimated coefficients are not statistically significant, they provide additional evidence that the school leadership professional development program may have had a modest impact on PLE test-scores.

Table 8: **Local average treatment effects on test-scores in 2019.**

Variable	Aggregate	Aggregate	English	English	Nonstem	Nonstem	Stem	Stem
Treated (LATE)	0.05	0.05	0.05	0.04	0.03	0.03	0.05	0.04
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
First-stage								
Treatment assignment	0.66***	0.66***	0.67***	0.66***	0.67***	0.66***	0.66***	0.66***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
H_0: weak instrument								
Robust F-statistic	168.03	172.05	175.74	175.36	174.02	175.203	167.35	171.06
Probability > F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Add. controls	No	Yes	No	Yes	No	Yes	No	Yes
ITT 2019	0.03	0.03	0.03	0.03	0.02	0.02	0.04	0.03
Number of students	30438	30438	30438	30438	30438	30438	30438	30438
Number of schools	276	276	276	276	276	276	276	276
R^2	0.18	0.18	0.15	0.15	0.17	0.18	0.16	0.16

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership, population density of the sector where the school is located in and the average grade per school in 2017. Additional controls refer to, whether a school was opened after 1994, electricity, total number of students and teachers, number of students taking the PLE exam in 2017, distance to the capital and nightlight activity. If information is missing for the additional controls, we replace the missing value with the average across all schools and include a dummy indicating whether a school has missing information. Standard errors are clustered at the school level. ITT records the intent-to-treat effect as estimated previously for 2019.

4.5 Placebo tests and alternative explanations

4.5.1 Placebo average treatment effects

We proceed by computing placebo average treatment effects. Table A4 shows placebo treatment effects for test-score results in 2017, 2016 and 2015. The estimated coefficients are small and insignificant implying again that the final sample was well balanced with regard to test-scores.

4.5.2 Placebo heterogenous treatment effects

Thus far, the evidence suggests that the SLPDP was more effective in public schools, which are also over-proportionately located in less densely populated areas, as well as in areas with less nightlight activity. However, these results could potentially be attributed to statistical chance. To address this issue, we do two robustness checks. First, we run the same regression with the interactions terms on test-scores from 2017 and 2016. Technically, we should not pick up any significant treatment effect.

In this regard, table A5 shows that none of the interaction terms is significant with regard to test-scores in 2017 and 2016. We also detect no significant difference between public and government-aided schools in the treatment group. Therefore, comparing the estimates in table A5 with the estimates in table 6 gives us some confidence that argue that the SLPDP had a modest impact on public schools.

Second, we conduct the same regression analysis with interaction terms on test-scores from 2018 and 2019, this time including cohort 2 in our estimations. Despite approximately 40 schools from this cohort having received some training by 2019, the inclusion of additional schools in the control group enhances the power and precision of the regressions. Encouragingly, the estimates remain consistent as shown in table A6. Also, particularly in relation to nightlight activity, the standard errors decrease, even though the coefficients remain insignificant.

4.5.3 Student enrollment

As previously outlined, another concern is the possibility that the program could have influenced weaker students to remain enrolled in P6 and take the Primary Leaving Examination (PLE). This scenario would result in having weaker students in the treatment group, while similar students may have dropped out in the control group and could, therefore, bias the average treatment effects down.

Table A7 displays the number of test-takers for public and government-aided schools based on treatment status between 2015 and 2019. There is limited evidence to suggest substantial differences in student numbers between treated and untreated government-aided schools over time. However, the table does reveal a potential increase of around 6 students in treated public schools compared to non-treated public schools. To address this issue, we conducted a regression analysis, regressing the number of students taking the PLE exam in 2018 or 2019 on the treatment group. We also include an interaction term between "treatment" and "government-aided schools," as well as district and school type fixed effects and the number of students taking the PLE exam in 2015, 2016, and 2017 per school.

As indicated in Table A8, the number of students taking the PLE exam in public schools increased by 3 in 2018 and 7 in 2019. Following the SLPDP student numbers may have increased by 6 in 2018 and decreased by 3 in 2019 in government-aided schools. However, none of the estimated coefficients is statistically significant. Given that the average number of students in P6 in public schools in 2017 was 119, an increase of 7 students would represent a six percentage point rise in the number of students taking the PLE exam. This estimate is relatively modest and would, therefore, not significantly bias our estimates downwards.

However, we conduct an additional robustness check by calculating Lee bounds for test-scores in 2018 and 2019.[22] Drawing upon our earlier estimations in table A8, we hypothesize that the SLPDP induced 7 students to take the P7 exam in treated schools. These students would have dropped out without the SLPDP.

Tables A9 and A10 present the lower and upper bounds of the treatment effects, along with their 95% confidence intervals, when excluding the 7 highest scoring or 7 lowest scoring students in each treatment school. These intervals therefore represent the range within the true treatment may fall, when considering students that would have been always observed, independently of treatment status. The confidence intervals encompass both negative and positive values. This suggest that a positive treatment effect may not be present overall and we can confidently state that our previous regressions do not underestimate the actual treatment effect (Tauchmann, 2014).

4.5.4 Spillover effects

One last concern is the potential for spill-over effects. In fact, treatment and control schools are allocated relatively close to each other with an average distance of only 2 km from one treatment school to one control school. To test for spillover effects, we restrict the sample to schools in cohort 3 and exclude those that received treatment in 2018 or 2019. We compute the distance of each school in cohort 3 that did not participate in the SLPDP with the closest school from cohort 1 or cohort 3 that participated in the SLPDP in 2018. We then run the following regression:

$$Y_{isd} = \alpha + \beta \times \text{Distance}_s + X_{isd} + \bar{Y}^{j}_{s,t-1} + Z_s + \lambda_d + \epsilon_{isd}, \tag{4}$$

where Y represents aggregate test-scores in 2018 or 2019 of individuals that went to a school s from cohort 3 in district d that did not participate in the SLPDP. The coefficient β captures the extent to which the distance of a school in cohort 3 to the closest school from cohort 1 or cohort 3 that participated in the SLPDP predicts test score gains. If β is negative and statistically significant, this suggests potential spillover effects, as schools that are close to treated schools exhibit higher student

[22] Lee bounds provide a range for the treatment effect rather than a single point estimate. Lee bounds discard either the highest or lowest values from the less attritted study arm and identify the ATE for those who would always be observed, regardless of their treatment status (Tauchmann, 2014). In this regard, Lee bounds are less extreme than simply replacing missing observations with the maximum or minimum values in the support of the outcome variable.

grades. We further include student characteristics X_{isd}, school characteristics Z_s, prior test-scores from 2017 $Y_{s,t-1}$, and district fixed effects λ_d. Standard errors are clustered at the school level. In this regard, table A11 shows that the distance to the closest school in cohort 1 that received treatment does not predict test-scores in 2018 or 2019 in cohort 3.

4.6 Cost-effectiveness analysis

While it is apparent that the program showed less efficacy in government-aided schools in terms of educational achievement, it remains valuable to undertake a cost-effectiveness analysis to assess the educational gains observed in public primary schools. This analysis, however, solely examines the impact on student test scores and does not encompass other potential outcomes, such as the well-being and mental health of students or teachers. These aspects, as suggested by Goldberg (2019), could also experience positive effects from SLPDPs. To compare the costs and benefits of the program, we follow the methodology proposed by Bhula et al. (2023); JPAL (2023); Walter (2020).

Based on the specifications, the school leadership professional training program has shown a significant increase in test-scores by 0.10 (including cohort 2) to 0.11 (without cohort 2) standard deviations in public schools. For simplification purposes, we opt to use an increase in test-scores by 0.10 standard deviations.

The number of students who took the PLE exam was 93 in 2018 and 106 in 2019 in public primary schools that participated in the SLPDP (table A7). However, the school leadership professional training program may have benefited not only students taking the PLE-exam, but the entire student population of a primary public school (Agirdag and Muijs, 2023; Anand et al., 2023; IIEP, 2023). Based on the 2018 Annual School Census data, the primary section of public primary schools that participated in the SLPDP in 2018 or 2019 had an enrollment of around 850 students.[23]

As previously mentioned, we only have PLE test-score data and are therefore unable to gauge the educational achievement of students in grades 1 to 5. To address this limitation, we assume a uniform increase in test scores across all class grades, approximating that every student in a public primary school may have experienced an improvement of approximately 0.11 standard deviations.[24]

The average cost of the program was USD 1412 in 2018 per participating headteacher comprising

[23] Please note that the Rwanda Ministry of Education reports an average of 1100 students enrolled in primary public schools across Rwanda in 2018, but does not provide a breakdown by district (NISR, 2018).

[24] It is worth noting that a potential criticism could be the uneven allocation of resources with schools potentially directing more resources to students in grade P6, who are preparing for the transition to secondary school. However, this factor remains unaccounted for in our analysis.

substantial and administrative costs.[25] To make these costs comparable to other interventions, researchers usually rely on the US-Dollar in 2011 as a baseline. Adjusting for the inflation rate implies that the overall cost per headteacher would equal USD 1265 in 2011. To take into account the cost spend on headteacher in government-aided schools, where the SLPDP did not have a measurable impact, we divide USD 1265 by the share of public schools given all public and government aided schools that received training: USD 1265 / (64/184) = USD 3637.

In a typical school of 850 students, cumulative learning gains would be 850 × 0.11 = 93.5 standard deviations. Translating these costs and benefits into total standard deviations of learning gains per USD 100, results into learning gains of 2.57 standard deviations per USD 100. This estimate places in the medium range of cost-effectiveness among education interventions with regard to test-scores, as reported by Kremer et al. (2013).

Recent critiques from the World Bank challenge the validity of exclusively assessing interventions based on their impact on learning gains (Angrist et al., 2020; Filmer et al., 2019). The conventional measure of standard deviations, often employed to quantify learning gains, computes improvements in relation to a localized distribution of test scores. This approach, however, poses challenges in evaluating whether the attained progress justifies the associated costs in absolute terms. Furthermore, policy makers may also be interested in evaluating whether improving learning or improving total schooling years is preferable. Therefore, we also apply a Learning-Adjusted Years of Schooling model (LAYS) that allows to compare the educational gains of the SLPDP with other interventions worldwide in terms of schooling years and learning gains against an absolute, cross-country standard.[26]

Assuming that the program led to an increase in test-scores by 0.11 standard deviations, when applying the benchmark of learning gains typically observed at approximately 0.80 standard deviations per school year in high-income countries, this translates to a Learning Adjusted Years of Schooling (LAYS) of 0.1375 per student. By multiplying 0.1375 with the total number of students in a primary school (850) and dividing the result by 36.37, we arrive at a LAYS of 3.21 per USD 100 spent. This estimate positions the intervention also in the medium effective range when using LAYS as a metric scale.[27]

Overall, this section underscores that the School Leadership Professional Development Program

[25] Substantial costs include the training sessions and the provision of learning materials. Administrative costs refer mainly to the registration fees and the examination fees of the candidates.

[26] The exact methodology of the LAYS approach is explained by the paper from Angrist et al. (2020).

[27] For comparison, please refer to figure 6, p.22, in Angrist et al. (2020).

(SLPDP) exhibits the potential to enhance learning outcomes at a commendably low cost, particularly when tailored to specific groups of schools or headteachers. While acknowledging some degree of uncertainty in our estimates, it is worth mentioning that more recent versions of the program operate with even lower costs.[28] This improvement is attributed to the reduction in administrative costs and the integration of online teaching modules. Moreover, registration costs, but in particular costs related to the examination of the headteachers can in theory be further reduced, as they are not substantial to the training itself.

5 Mechanisms

This section aims to understand the observed positive treatment effects for public schools as compared to government-aided schools. Public schools are also more commonly found in less urban areas and in areas with less nightlight activity. First, we present summary statistics on headteachers' characteristics collected by VVOB. Regrettably, this data solely encompasses headteachers actively involved in the program, lacking information on school principals who did not participate. Consequently, we are unable to use headteacher characteristics to estimate any causal effects on student test-scores. Second, we leverage interviews to comprehend how the SLPDP changed school management within schools.

5.1 Characteristics of headteachers participating in the SLPDP

VVOB collected information on the gender, the birth year, the education of the headteacher and the amount of years the headteacher is already in charge in a given school. Table 9 shows descriptive statistics on headteachers in both public and government-aided schools that participated in the SLPDP and are from cohort 1, 2 or 3.

Some disparities may exist between headteachers in public and government-aided schools, particularly concerning age and gender. In government-aided schools, the average proportion of female headteachers is 23%, contrasting with only 9%, in public schools. In addition, headteachers in public primary schools are, on average, 2 years younger than their counterparts in government-aided schools. Despite these variations, no discernible differences are observed in educational attainment,

[28] A total of 650 recently appointed headteachers in primary schools throughout Rwanda are undergoing training in four cohorts spanning from 2023 to 2026. It is crucial to note that the assignment of individuals across these cohorts is not assured to be random. Consequently, this lack of randomness implies that evaluating these four cohorts using a conventional Randomized Controlled Trial (RCT) setup is not feasible.

such as possessing a post-secondary education degree versus completing only secondary school or in the total years of serving as a headteacher between public and government-aided schools.

Table 9: **Characteristics of headteachers in public and government-aided schools.**

Variable	Sample mean	Public	Gov-aided	Mean diff.	P-value	Size public	Size gov-aided
Birth year	1975.02	1973.69	1975.73	2.63*	0.07	64	120
	(8.54)	(9.05)	(8.20)				
Female headteacher	0.18	0.09	0.23	0.10	0.07	64	120
	(0.39)	(0.29)	(0.42)				
Experience in school	6.70	6.66	6.72	-0.52	0.62	64	120
	(6.18)	(6.80)	(5.85)				
Education	0.55	0.61	0.52	-0.05	0.46	64	120
	(0.50)	(0.49)	(0.50)				
F-test of joint significance	F-value: 2.46	P-value: 0.05**					

NOTE: Significance levels: *<0.10; **<0.05 ***<0.01. The values in brackets denote standard deviations. The mean difference is computed by regressing treatment on the covariate including district fixed effects and the schooltype. Standard errors are clustered at the school level. There 64 public and 120 government aided schools in the sample. Education is a dummy and is one if a headteacher has received any training beyond secondary school, such as post-secondary training courses or an university degree. Experience in school refers to the total number of years the headteacher is in charge in the given school.

To assess whether headteacher characteristics are predictive of aggregate test-scores among students, we estimate the following equation:

$$Aggregate_{isdt} = \alpha + \beta_1 \times Treated_s^{2018} + \beta_2 \times Treated_s^{2018} \times Charact_s + \bar{Y}_{s,t-1}^{j} + Z_s + X_i + \lambda_t + \delta_d + \epsilon_{isdt}, \quad (5)$$

where "aggregate" refers to the aggregate test-scores of student i attending school s in district d and taking the PLE exam in year t (either 2018 or 2019). The coefficient of interest is β_1, capturing the impact of a headteacher in school s who participated in the SLPDP in 2018 compared to 2019. We consider headteachers from cohort 1, 2 or 3.

In this context, β_2 represents the differential impact of the SLPDP concerning headteacher characteristics, such as gender, leading a government-aided school, education level and the duration the headteacher has been in charge in the given school. We control for average student grades in 2017 $\bar{Y}_{s,t-1}^{j}$, student characteristics X_i, school characteristics Z_s, district fixed effects δ_d and test-year fixed effects λ_t. Standard errors are clustered at the school level.

Table 10 consistently supports our earlier results. Specifically, for headteachers in public schools that received training in 2018 versus those that received partial training in 2019, there is a noteworthy 0.12

standard deviation increase in test-scores at the ten percent significance level. The overall impact on headteachers from government-aided schools is nearly negligible. Furthermore, factors such as the headteacher's age (column 2), education level (column 4) and time in charge show no discernible influence on the observed effects. In particular, the findings regarding the education level of the headteacher align with existing literature, which consistently indicates that the education of school leaders often does not account for significant variations in teacher quality (Aslam and Kingdon, 2011; Crawfurd and Rolleston, 2020; Metzler and Woessmann, 2012).

Table 10: **Heterogeneous treatment effects between headteachers that participated in the training program in cohort 1 or 2 on test-scores in 2018 and 2019.**

Variable	(1) Aggregate	(2) Aggregate	(3) Aggregate	(4) Aggregate	(5) Aggregate
Treated	0.12	-0.28	0.07	0.02	0.01
	(0.07)	(9.19)	(0.05)	(0.07)	(0.06)
Government aided	0.03				
	(0.07)				
Government aided x treated	-0.11				
	(0.09)				
Birth year headteacher		0.00			
		(0.00)			
Birth year headteacher × treated		-0.00			
		(0.00)			
Female headteacher			0.09		
			(0.07)		
Female headteacher × treated			-0.10		
			(0.09)		
Post-secondary				0.04	
				(0.06)	
Post-secondary × treated				0.03	
				(0.09)	
Experience					0.00
					(0.00)
Experience × treated					(0.01)
					(0.01)
Test year fe	yes	yes	yes	yes	yes
Number of students	36189	36189	36189	36189	36189
Number of schools	184	184	184	184	184
R^2	0.19	0.19	0.19	0.19	0.19

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. Reported are aggregated test-scores in columns (1)-(4), stem test-scores in columns (5)-(8) and non-stem test-scores in columns (9)-(12). The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership, population density in the given sector and the average grade per school in 2017. Standard errors are clustered at the school level. Reported are unadjusted p-values** and adjusted p-values$_{*}$ for multiple hypothesis testing. All outcomes variables in the table are grouped as one family.

Moreover, we observe 0.07 standard deviation increase in test-scores for male headteachers, as indicated in column 4. This effect is slightly insignificant, but in contrast, there is no effect for female headteachers. This could be attributed to the challenges women face in attaining headteacher positions, potentially resulting in a pool of female headteachers who already possess higher management capabilities. The coefficient for female headteachers is positive, with a 0.11 standard deviation, suggesting that students in schools led by female headteachers perform better. However, it is also plausible that women encounter greater challenges in implementing changes within schools. This aligns with existing literature on gender differences in leadership style and gender stereotypes (Alan et al., 2020; Bandiera et al., 2021; Ertac and Gurdal, 2019). We propose conducting further research into gender-specific dynamics related to school management and SLPDPs.

5.2 Management practices

The analysis of headteacher characteristics does not account for a significant portion of the observed performance disparities between public and government-aided schools. However, quantitative evidence indicates that students in government-aided schools, which are predominantly situated in urban areas, initially perform better than those in public schools. This suggests that government-aided schools benefit from a stronger student base from the start, potentially leading to inherently better school performance. Therefore, the impact of the SLPDP may be less pronounced in these schools, whereas in public schools, the program has notably encouraged headteachers to enhance their educational practices.

In the absence of quantitative data regarding the behaviors of principals and teachers, this section presents findings from qualitative interviews conducted by VVOB on primary headteachers who participated in the training program in either 2018 or 2019 (Stones, 2021; Mukingambeho et al., 2023). In total, 49 primary headteachers, including 4 female headteachers, as well as 24 math teachers in all six targeted districts, were interviewed. We proceed by highlighting the positive effects of the SLPDP on school leader management practices before discussing the challenges that may explain why educational gains are stronger in public compared to government-aided schools.

5.2.1 Positive effects of the SLPDP on the headteacher

Headteachers reported that the program helped to develop an understanding of school management and school strategy. The new knowledge bolstered their confidence and satisfaction, fostering a more collaborative approach to decision-making. For instance, one school principal reported: *"The part of the course that stuck with me was leadership style. It helped me work better with the teachers in a way that brought success to student performance and the teachers started to enjoy their work."*

Moreover, headteachers became more hands-on and actively participated in teaching activities, providing essential support and training for teachers. In this regard, school principals reported a better understanding of the importance of a structured on-boarding process for teachers. A headteacher's perspective illustrates this shift: *"Before, we would send a teacher to the classroom with only teaching materials and without induction. Today, we recognize that teachers should receive proper induction. It is also important to emphasize the need to implement the curriculum correctly."*

Last, the SLPDP introduced the significance of stakeholders' involvement in school management. Headteachers started to recognize the value of collaboration with parents and community members. For instance, regular parent meetings became a norm in 75% of the interviews, reflecting a proactive engagement approach.

5.2.2 Challenges for the implementation of the SLPDP

However, headteachers also report certain challenges that hinder the effectiveness of implementing the SLPDP. One of the major issues is time constraints. Principals, burdened with multiple responsibilities, struggle to fully implement the acquired knowledge and to support teachers adequately. The scarcity of time hampers their ability to coach and guide new teachers effectively. While some teachers experience increased assistance and a conducive work environment, others find themselves lacking essential feedback and guidance. One teacher pointed out that *"the headmaster can only provide teaching materials upon availability and overall support is still insufficient."*, while another teacher stated that *"the principal is often busy with other tasks and rarely has time to help new teachers. Their schedule does not include any time to coach or assist a new teacher."*

Time constraints are a more significant challenge in government-aided schools due to distinct ownership structures. As mentioned, public schools are entirely operated by the government, while government-aided schools are often owned by churches, but receive support from the government. Therefore, headteachers from government-aided schools not only need to report to the government, but also to a broader set of stakeholders compared to their public counterparts. These extensive responsibilities outside the school setting limit their on-site presence. Much of their time is consumed by administrative duties, meetings and councils at sector or district levels, further burdening their overloaded schedules. A headteacher from a government-aided school highlights their multifaceted role: *"In primary schools, headteachers juggle various responsibilities, acting as leaders, secretaries, bursars, heads of studies, discipline overseers, Information Technology officers and liaisons with the community and local administration offices."*

Overall, headteachers in government-aided schools may need to involve more stakeholders, such as the school owners or the government, when requesting school management changes following the SLPDP. The ownership structure, therefore, adds an extra layer of complexity to their responsibilities, which potentially hinders their ability to focus on on-site improvements.

6 Conclusion

This study evaluates the effects of a large School Leadership Professional Development (SLPDP) program on human capital development in Rwanda. Although no significant average treatment effects on student test scores are found, the program increased test scores in public schools compared to government-aided schools by approximately 0.10 to 0.11 standard deviations. This increase is considered a small effect in the context of education interventions and their impact on test scores (Bhula et al., 2023). While the effect may be small, as outlined, the program has the potential to increase student test scores in a cost-effective manner, as usually only one headteacher per school is trained.

Moreover, we demonstrate that school leadership professional development programs need to target specific schools or headteachers in charge. Our program proved to be more effective in public schools compared to government-aided schools. We attribute this to three distinct patterns: first, public schools generally face challenges such as limited resources, weaker management quality, and lower initial student performance. This context suggests that the School Leadership and Development Program (SLPDP) is likely to have a more pronounced beneficial impact on schools that start with these disadvantages, improving both management practices and student outcomes. Second, if headteachers lack the time to effectively implement changes in school management proposed by the SLPDP, such programs might not succeed in the future. Third, ownership structure may also diminish the effects of SLPDP. Having more stakeholders could imply that fewer changes can be implemented. Overall, the SLPDP may be more effective when targeting headteachers with initial weaknesses, when headteachers can overcome time constraints and also when headteachers are faced with a less complex decision-making, when introducing meaningful changes.

There are two main limitations of this study. First, we solely focus on student test- scores; future research should explore the impact of school leader training on other dimensions of student well-being. Second, we examine the effect of the program in the short-term, i.e., 1-2 years after its implementation. It is worth noting that the impact of the school leadership professional development program may manifest more prominently in the long run. Consequently, the prospect of follow-up studies becomes intriguing for a comprehensive understanding of its enduring effects.

Future research should also focus on enhancing the effectiveness of school leadership development programs by improving their various components and assessing their cost-effectiveness. Additionally, it would be interesting to investigate whether the school leadership training program impacts students differently depending on their grade level. Specifically, understanding how resources are allocated towards students, whether at the end or the beginning of primary education, could provide valuable insights.

References

Agirdag, Orhan and Daniel Muijs, "School leadership development and academic achievement: Effectiveness of the High Performing Schools programme," *International Journal of Educational Research*, 2023, *122*.

Alan, Sule, Seda Ertac, Elif Kubilay, and Gyongyi Loranth, "Understanding Gender Differences in Leadership," *The Economic Journal*, February 2020, *130* (626), 263–289.

Anand, Gautam, Aishwarya Atluri, Lee Crawfurd, Todd Pugatch, and Ketki Sheth, "Improving school management in low and middle income countries: A systematic review," *Working Paper*, 2023.

Angrist, Noam, David K. Evans, Deon Filmer, Rachel Glennerster, F. Halsey Rogers, and Shwetlena Sabarwal, *How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions using the New Learning-Adjusted Years of Schooling Metric* Policy Research Working Papers, The World Bank, October 2020.

Aslam, Monazza and Geeta Kingdon, "What can teachers do to raise pupil achievement?," *Economics of Education Review*, June 2011, *30* (3), 559–574.

Aturupane, Harsha, Paul Glewwe, Tomoko Utsumi, Suzanne Wisniewski, and Mari Shojo, "The impact of Sri lanka's school-based management programme on teachers' pedagogical practices and student learning: evidence from a randomised controlled trial," *Journal of Development Effectiveness*, 2022, *14* (4), 285–305.

Bandiera, Oriana, Greg Fischer, Andrea Prat, and Erina Ytsma, "Do Women Respond Less to Performance Pay? Building Evidence from Multiple Experiments," *American Economic Review: Insights*, December 2021, *3* (4), 435–454.

Benjamini, Yoav and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1995, *57* (1), 289–300.

Bhula, Radhika, Meghan Mahoney, and Kyle Murphy, "Conducting cost-effectiveness analysis (CEA)," *JPAL*, 2023.

Blimpo, Moussa P., David Evans, and Nathalie Lahire, "Parental Human Capital and Effective School Management: Evidence from The Gambia," 2015. Publisher: World Bank, Washington, DC.

Bloom, Nicholas, Renata Lemos, and Raffaela Sadun, "Does Management Matter in schools?," *The Economic Journal*, 2015, *125* (584).

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin, "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals," *NBER Working Paper Series*, 2012, *17803*.

Bruederle, Anna and Roland Hodler, "Nighttime lights as a proxy for human development at the local level," *PLOS ONE*, 2018, *13* (9).

Crawfurd, Lee, "School Management and Public–Private Partnerships in Uganda," *Journal of African Economies*, November 2017, *26* (5), 539–560.

_ , "Accounting for repetition and dropout in contemporaneous cross-section learning profiles: Evidence from Rwanda," *International Journal of Educational Development*, September 2021, *85*, 102443.

_ **and Caine Rolleston**, "Long-run effects of teachers in developing countries," *Review of Development Economics*, 2020, *24* (4), 1279–1299. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/rode.12717.

de Hoyos, Rafael, Alejandro J Ganimian, and Peter A Holland, "Great Things Come to Those Who Wait: Experimental Evidence on Performance-Management Tools and Training in Public Schools in Argentina," Technical Report 2020.

_ **,** _ **, and** _ , "Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity Building for Public Schools in Argentina," *The World Bank Economic Review*, May 2021, *35* (2), 499–520.

de Ricardo Paes Barros, de Mirela Carvalho, Samuel Franco, Beatriz Garcia, Ricardo Henriques, and Laura Machado, "Assessment of the Impact of the Jovem de Futuro Program on Learning," 2019.

Ertac, Seda and Mehmet Y. Gurdal, "Preference Communication and leadership in group decision-making," *Journal of Behavioral and Experimental Economics*, June 2019, *80*, 130–140.

et al. Goldberg, "Effectiveness of interventions adopting a whole school approach to enhancing social and emotional development: a meta-analysis," *European Journal of Psychology of Education*, 2019, *34* (4), 755–782.

Filmer, Deon, Halsey Rogers, Noam Angrist, and Shwetlena Sabarwal, "Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education," *Policy Research Working Paper*, 2019, *8591*.

Grissom, Jason A., Anna J. Egalite, and Constance A. Lindsay, "How Principals Affect Students and Schools: A Systematic Synthesis of Two Decades of Research," Technical Report, Wallace Foundation February 2021.

IIEP, "School leadership," *UNESCO*, 2023.

JPAL, "Conducting cost-effectiveness analysis (CEA)," *The Abdul Latif Jameel Poverty Action Lab (J-PAL)*, 2023.

King, Elisabeth, ed., *Schooling after Genocide*, Cambridge: Cambridge University Press, 2013.

Kremer, Michael, Conner Brannen, and Rachel Glennerster, "The challenge of education and learning in the developing world," *Science*, 2013, *340* (6130), 297–300. Publisher: American Association for the Advancement of Science.

Lassibille, Gérard, "Improving the management style of school principals: results from a randomized trial," *Education Economics*, 2016, *0* (0), 1–21.

Leaders, Global School, "Evidence Review Report: A Review of the Empirical Research on School Leadership in the Global South," Technical Report, Global School Leaders 2020.

Leithwood, Kenneth, Alma Harris, and David Hopkins, "Seven strong claims about successful school leadership revisited," *School Leadership & Management*, 2020, *40* (1), 5–22.

Lemos, Renata, Karthik Muralidharan, and Daniela Scur, "Personnel Management and School Productivity: Evidence from India," Technical Report w28336, National Bureau of Economic Research, Cambridge, MA January 2021.

Lopez, Ann and Peter Rugano, "Educational Leadership in Post-Colonial Contexts: What Can We Learn from the Experiences of Three Female Principals in Kenyan Secondary Schools?," *Education Sciences*, 2018, *8* (3), 99.

McCord, Gordon Carlos and Mario Rodriguez-Heredia, "Nightlights and Subnational Economic Activity: Estimating Departmental GDP in Paraguay," *Remote Sensing*, February 2022, *14* (5), 1150.

Metzler, Johannes and Ludger Woessmann, "The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation," *Journal of Development Economics*, November 2012, *99* (2), 486–496.

Miller, Ashley, "Principal turnover and student achievement," *Economics of Education Review*, October 2013, *36*, 60–72.

MOE, "Education sector strategic plan. 2018-2024. Rwanda | Documents | Global Partnership for Education," 2018.

Mukingambeho, Delphine, Heredion Ndayisabye, Gabriel Nizeyimana, Jean Leonard Buhigiro, and Dan Imaniriho, "Exploring the outcomes of the Continuous Professional Development diploma programme in effective school leadership in Rwanda," *Rwandan Journal of Education*, 2023, *7* (1), 72–89. Number: 1.

Muralidharan, Karthik, "Field experiments in education in developing countries," in "Handbook of economic field experiments," Vol. 2, Elsevier, 2017, pp. 323–385.

_ **and Abhijeet Singh**, "Improving Public Sector Management at Scale? Experimental Evidence on School Governance in India," *Nber Working Paper 28129*, 2020.

NISR, "Statistical YearBook 2018 | National Institute of Statistics Rwanda," 2018.

_ , "Home | National Institute of Statistics Rwanda," 2024.

Nzaramba, Samuel, Carl Menon, and Phil Leonard, "Gender gaps in secondary schools in Rwanda: Evidence from national assessments," Report, University of Cambridge, Cambridge 2021.

Patrinos, Harry Anthony, Felipe Barrera-Osorio, and Juliana Guaqueta, "The Role and Impact of Public-Private Partnerships in Education," Technical Report, World Bank, Washington, DC 2009.

REB, "Professional standards for effective school leadership," 2020.

Romero, Mauricio, Juan Bedoya, Monica Yanez-Pagans, Marcela Silveyra, and Rafael De Hoyos, "Direct vs indirect management training: Experimental evidence from schools in Mexico," *Journal of Development Economics*, January 2022, *154.*

Saux, Laura Le, Anthony Onwuegbuzie, Philip Leonard, Stoelinga, Dimitri, and Sabates, Ricardo, "School leadership and school quality in secondary education in Rwanda | Unesco IIEP Learning Portal," Technical Report, UNESCO 2021.

Scheunpflug, Annette, Mark Wenz, Mimii Brown Rubindamayugi, Jean Kasereka Lutswamba, Frederick Njobati, Christine Nyiramana, Samuel Mutabazi, Claude Ernest Njoya, Onja Raharijaona, and Quentin Wodon, "Relationships between Christian schools and the state: a comparative analysis for five sub-Saharan African countries," *International Studies in Catholic Education*, July 2021, *13* (2), 163–174.

Stones, Three, "End Evaluation of the programme Mathematics Achievements in Rwandan Schools – Girls on MARS 2017-2021 Rwanda," Technical Report 2021.

Tauchmann, Harald, "Lee (2009) Treatment-Effect Bounds for Nonrandom Sample Selection," *The Stata Journal: Promoting communications on statistics and Stata*, 2014, *14* (4), 884–894.

Trines, Stefan, "Education in Rwanda," Technical Report, WENR 2019.

UNESCO, "Education: From disruption to recovery," March 2020.

UNESCO, "Non-state actors in education in Rwanda," 2023.

UNICEF, "Situation of children in Rwanda | UNICEF Rwanda," November 2024.

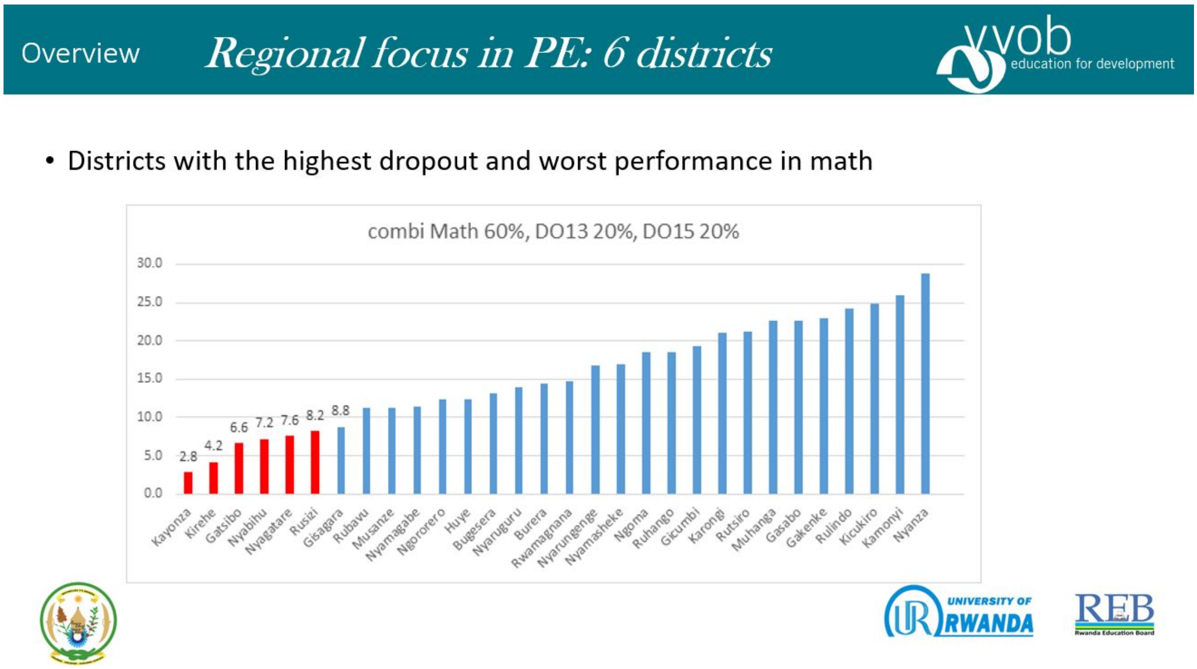
Walter, Torsten Figueiredo, "Misallocation in the Public Sector? Cross-Country Evidence from Two

Million Primary Schools," *STICERD Working Paper 70*, 2020.

World Bank, *Learning to Realize Education's Promise. World Development Report 2018,* Washington DC: World Bank, 2018.

Appendix

Figure A1: District selection



NOTE: This figure shows the ranking of districts. Districts were ranked from lowest to highest with regard to their math performance in P6 and the school drop out rate in P6 in 2013 and 2016. The ranking was weighted with 60 percent for the math performance and 20 percent each for the school drop-out rate in 2013 and 2016. The lowest six districts were picked for implementation.

Table A1: **Randomisation across cohorts**

	Cohort 1	**Cohort 2**	**Cohort 3**
School-type			
Only primary	63%	68%	66%
9 YBE	15%	13%	13%
12 YBE	22%	19%	21%
Total	100%	100%	100%
School-ownership			
Government aided	50%	54%	52%
Private	15%	16 %	14%
Public	35%	29%	33%
Total	100%	100%	100%
Districts			
Gatsibo	32	32	32
Kayonza	30	30	30
Kirehe	18	18	18
Nyabihu	27	27	27
Nyagatare	34	34	34
Rusizi	34	34	34
Total	175	175	175

Table A2: **Description of variables across cohorts**

Variable	**Description**	**Source**
P6 classrooms 2017	Number of classrooms available in a school in 2017 for P6 students	ASC 2017
Total students 2017	Number of students in a school in 2017	ASC 2017
Students in P6 2017	Number of students in P6 in a school in 2017	ASC 2017
PLE Takers	Number of students sitting PLE examination	Test-score data
Teachers 2017	Number of teachers in a given school in 2017	ASC 2017
Establishment date	Year the school was opened	ASC 2018
Electricity 2017	Availability of electricity in a school	ASC 2018
Water 2017	Availability of water in a school	ASC 2018
School feeding program	Free school meals in a school	ASC 2017
Gross Cell product 2005	Regional equivalent of gross domestic product	Prio-Grid
Urban	Schools in urban location	Test-score data
Night light activity 2012	Derived from the average visible band digital number (DN) of cloud-free light detection's multiplied by the percent frequency of light detection	Prio Grid
Pop density 2012	Population density per $1km^2$ per sector	Population Census 2012

NOTE: ASC refers to Annual School Census data. Test-score data refers to the test-score in our possession between 2015 and 2019. Prio-Grid is data that is public available (Patrinos et al., 2009) and population census to the population census data (NISR, 2024)

Table A3: **School inputs after school-ownership**

Variable	Mean public	Mean gov-aided	Difference	Number public	Number gov-aided
Electricity	0.46	0.53	-0.07	108	168
Water	0.55	0.55	0.00	108	168
School feeding	0.39	0.24	0.15	108	168
Social Science books per student	0.05	0.05	0.00	27	44
Math books	0.06	0.06	0.00	27	44
English book	0.07	0.16	-0.09	26	43
Kinyarwanda books	0.06	0.08	-0.02	25	39
Laptops per teacher	0.03	0.11	-0.08	63	85
Laptops per student	0.09	0.07	0.02	63	85
Student to teacher ratio	95	125	-30	108	168
Student to classroom ratio	51	48	3	108	168

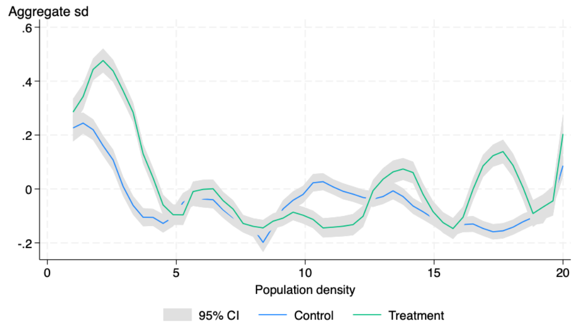
NOTE: This table was computed with the information from the Annual School Census data from 2017. For some variable information is missing being the reason why the student numbers are not constant with regard to the different variables.

Table A4: **Placebo treatment effects on test-scores in 2017, 2016 and 2015.**

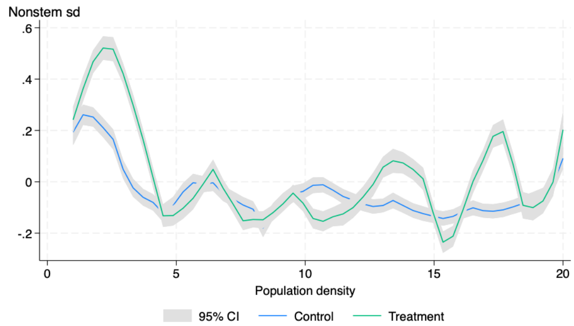
Variable	(1) Aggregate	(2) Nonstem	(3) Stem	(4) Aggregate	(5) Nonstem	(6) Stem	(7) Aggregate	(8) Nonstem	(9) Stem
Treatment	0.06	0.05	0.03	0.01	0.01	-0.01	0.04	0.06	0.01
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Exam year	2017	2017	2017	2016	2016	2016	2015	2015	2015
Number of students	24557	24557	24557	20320	20320	20320	17671	17671	17671
Number of schools	275	275	275	275	275	275	271	271	271
R^2	0.13	0.13	0.11	0.13	0.13	0.11	0.12	0.12	0.12

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, the birth year of the student, gender of the student, school-type, school-ownership, population density in the given sector and the number of test-takers in the school during the test-year. Standard errors are clustered at the school level.

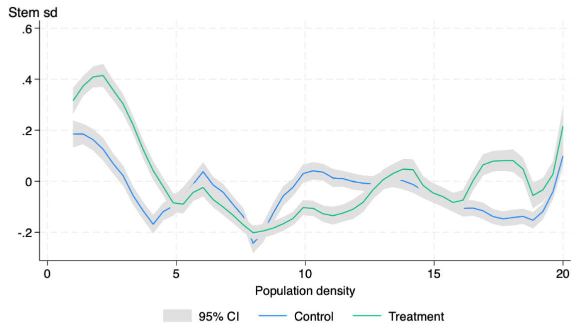
Figure A2: Local polynomial plots of test score residuals and population density, by subject



(a) Aggregate grade



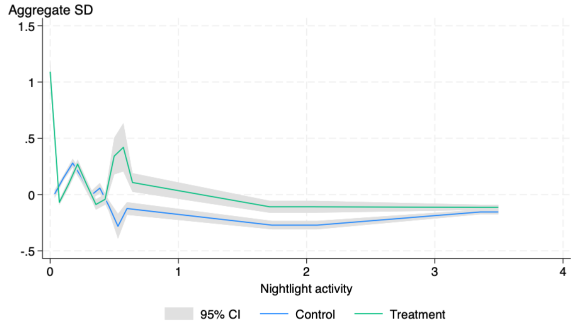
(b) Nonstem grade



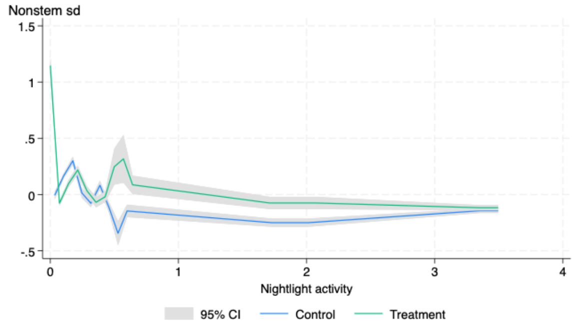
(c) Stem grade

NOTE: Line plots with a quadratic polynomial fit (kernel epanechnikoz) of degree 2 for both the control and treatment group that examine the relationship between population density (divided into 20 quintiles) and the test-score grades (measured in standard deviations). The local smooth is at measured for 50 observations.

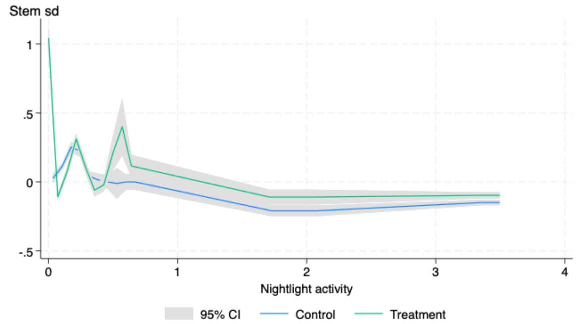
Figure A3: Local polynomial plots of test score residuals and nightlight activity, by subject



(a) Aggregate grade



(b) Nonstem grade



(c) Stem grade

NOTE: Line plots with a quadratic polynomial fit (kernel epanechnikoz) of degree 2 for both the control and treatment group that examine the relationship between nightlight activity and the test-score grades (measured in standard deviations). The local smooth is at measured for 50 observations.

Table A5: **Placebo heterogenous treatment effects on test-scores in 2017 and 2016.**

Variable	(1) Ag.	(2) Ag.	(3) Ag.	(4) Ag.	(5) Stem	(6) Stem	(7) Stem	(8) Stem	(9) Non	(10) Non	(11) Non	(12) Non
Treatment	0.00	0.03	-0.02	-0.03	0.01	0.02	-0.03	-0.05	-0.01	0.01	-0.05	-0.03
	(0.04)	(0.06)	(0.06)	(0.06)	(0.04)	(0.06)	(0.06)	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)
Female student	-0.36^{***}_{***}				-0.38^{***}_{***}				-0.39^{***}_{***}			
	(0.02)				(0.02)				(0.02)			
Treat. × Female student	0.02				0.01				0.02			
	(0.03)				(0.03)				(0.03)			
Government aid.		-0.00				-0.01				-0.02		
		(0.06)				(0.06)				(0.06)		
Treat. × Government aid.		-0.02				-0.01				-0.01		
		(0.08)				(0.08)				(0.08)		
Nightlight ≥ Med.			0.05				0.02				0.01	
			(0.07)				(0.07)				(0.07)	
Treat. × Nightlight ≥ Med.			0.07				0.09				0.10	
			(0.07)				(0.07)				(0.07)	
Pop density ≥ Med.				-0.13^{*}				-0.15^{**}				-0.08
				(0.07)				(0.06)				(0.07)
Treat. × Pop density ≥ Med.				0.07				0.10				0.04
				(0.07)				(0.07)				(0.08)
Test year fe	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of students	44877	44877	44877	44877	44877	44877	44877	44877	44877	44877	44877	44877
Number of schools	276	276	276	276	276	276	276	276	276	276	276	276
R^2	0.20	0.20	0.20	0.20	0.18	0.18	0.18	0.18	0.22	0.22	0.22	0.22

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. Reported are aggregated test-scores in columns (1)-(4), stem test-scores in columns (5)-(8) and non-stem test-scores in columns (9)-(12). The standard regression includes district fixed effects, the birth year of the student, the gender of the student, the population density per sector in 2012, school-type, school-ownership and the average grade per school in the given subject in 2015. Further included are test-year fixed effects and the number of test-takers in 2016 and 2017. Standard errors are clustered at the school level. Reported are unadjusted p-values** and adjusted p-values$_{*}$ for multiple hypothesis testing. All outcomes variables in the table are grouped as one family.

Table A6: **Heterogeneous treatment effects on test-scores in 2018 and 2019 including cohort 2.**

Variable	(1) Ag.	(2) Ag.	(3) Ag.	(4) Ag.	(5) Stem	(6) Stem	(7) Stem	(8) Stem	(9) Non	(10) Non	(11) Non	(12) Non
Treatment	0.03	0.10**	0.06	0.04	0.04	0.08**	0.06	0.05	0.02	0.09	0.05	0.01
	(0.04)	(0.05)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Female student	-0.21$^{***}_{***}$				-0.26$^{***}_{***}$				-0.23$^{***}_{***}$			
	(0.02)				(0.01)				(0.01)			
Treat. × Female student	-0.02				-0.04*				-0.01			
	(0.02)				(0.02)				(0.02)			
Government aid.		0.01				0.00				0.01		
		(0.04)				(0.04)				(0.04)		
Treat. × Government aid.		-0.14**				-0.11*				-0.13*		
		(0.06)				(0.06)				(0.07)		
Nightlight ≥ Med.			0.03				0.04				0.03	
			(0.05)				(0.05)				(0.06)	
Treat. × Nightlight ≥ Med.			-0.09				-0.09				-0.08	
			(0.06)				(0.05)				(0.06)	
Pop density ≥ Med.				0.01				0.03				-0.05
				(0.04)				(0.04)				(0.04)
Treat. × Pop density ≥ Med.				-0.05				-0.08				-0.01
				(0.06)				(0.05)				(0.06)
Test year fe	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of students	84880	84880	84880	84880	84880	84880	84880	84880	84880	84880	84880	84880
Number of schools	412	412	412	412	412	412	412	412	412	412	412	412
R^2	0.20	0.20	0.20	0.20	0.17	0.17	0.17	0.17	0.19	0.20	0.19	0.19

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. Reported are aggregated test-scores in columns (1)-(4), stem test-scores in columns (5)-(8) and non-stem test-scores in columns (9)-(12). The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership and the average grade per school in the given subject in 2017. Standard errors are clustered at the school level. Reported are unadjusted p-values** and adjusted p-values$_{*}$ for multiple hypothesis testing. All outcomes variables in the table are grouped as one family.

Table A7: **Number of PLE test-takers according to test-year.**

Variable	PLE-Takers 2015	PLE-Takers 2016	PLE-Takers 2017	PLE-Takers 2018	PLE-Takers 2019
Public, Treatment = 1	66	74	84	93	106
Public, Treatment = 0	69	84	102	105	112
Difference	-3	-10	-18	-12	-6
Change	-	-7	-8	+6	+6
Number of schools	108	108	108	108	108
Gov-aided, Treatment = 1	59	67	80	93	105
Gov-aided, Treatment = 0	68	74	93	99	117
Difference	-9	-7	-13	-6	-12
Change	-	+2	-6	+7	-6
Number of schools	168	168	168	168	168

NOTE: This table summarizes the number of PLE test-takers after treatment status and exam year.

To test whether the SLPDP increased the number of students taking the PLE exam in 2018 or 2019, we run the following regression:

$$Testtakers_s^t = \alpha + \beta Treatment_s + X_s + \delta_d + \epsilon_s, \tag{6}$$

where $Testtakers_s^t$ refers to the number of students taking the PLE exam in 2018 or 2019. $Treatment_s$ is a dummy indicating, whether a school belonged to the first treatment group. X_s are school level control variables as specified in the table notes and δ_d are district fixed effects.

Table A8: **Program effect on the number of test-takers in 2018 and 2019.**

Variable	PLE-Takers 2018	PLE-Takers 2018	PLE-Takers 2019	PLE-Takers 2019
Treatment	4.54	2.83	1.01	7.14
	(2.90)	(4.46)	(3.87)	(6.36)
Treated x Gov aided		2.79		-10.01
		(5.88)		(8.37)
Number of schools	276	276	276	276
R^2	0.87	0.87	0.77	0.77

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Treatment refers to schools from the first cohort. Treatment effects and standard errors are reported. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, schooltype and school-ownership, as well as the number of students taken the PLE exam in 2017, 2016 and 2015. Standard errors are robust.

Table A9: Leebounds for test-scores in 2018.

Subject	Lower and upper bound
Aggregate	[-0.07***, 0.20***]
	(-0.09, 0.22)
English	[-0.10***, 0.16***]
	(-0.12, 0.19)
Nonstem	[-0.05***, 0.16***]
	(-0.07, 0.18)
Stem	[-0.11***, 0.12***]
	(-0.13, 0.13)

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. The sample consists of 27575 and 26735 selected observations. The second columns reports estimated Lee bounds (first row) and the 95% confidence region for the identified set (second row). The lee-bounds are tightened by the gender of respondent and two dummies indicating if the school is a government aided school and if the school offers primary or primary and secondary education.

Table A10: Leebounds for test-scores in 2019.

Subject	**Lower and upper bound**
Aggregate	[-0.06***, 0.19***]
	(-0.08, 0.21)
English	[-0.08***, 0.39***]
	(-0.10, 0.41)
Nonstem	[-0.06***, 0.06***]
	(-0.08, 0.08)
Stem	[-0.11***, 0.17***]
	(-0.13, 0.19)

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. The sample consists of 31142 and 30302 selected observations. The second columns reports estimated Lee bounds (first row) and the 95% confidence region for the identified set (second row). The lee-bounds are tightened by the gender of respondent and two dummies indicating if the school is a government aided school and if the school offers primary or primary and secondary education.

Table A11: **Spillover effects on schools in cohort 3.**

Variable	(1) Aggregate 2018	(2) Aggregate 2018	(3) Aggregate 2019	(4) Aggregate 2019
Distance	0.01	0.01	0.00	-0.00
	(0.01)	(0.02)	(0.02)	(0.02)
Add. controls	no	yes	no	yes
Number of students	10148	10148	11507	11507
Number of schools	97	97	97	97
R^2	0.20	0.20	0.17	0.17

NOTE: Levels of significance: *<0.10; **<0.05 ***<0.01. Distance refers to the closest distance of a school in cohort 3 to any school that participated in the SLPDP in 2018. Test-scores are standardized with mean 0 and standard deviation 1. The standard regression includes district fixed effects, the birth year of the student, the gender of the student, school-type, school-ownership, population density and the average grade per school in 2017. Additional controls refer to the total number of students and teachers, night light activity within the grid of the school, a dummy indicating if a school was opened after 1996, distance to the capital, population density in the given sector and whether the school has electricity. If information is missing for the additional controls, we replace the missing value with the average across all schools and include a dummy indicating whether a school has missing information. Standard errors are clustered at the school level.