Journal of Development Economics 177 (2025) 103545

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Improving school leadership in Rwanda

Simeon Lauterbach [a,*], Lee Crawfurd [b], Jocelyne C. Kirezi [c], Aimable Nsabimana [d], Jef Peeraer [c]

[a] Geneva Graduate Institute, Switzerland
[b] Center for Global Development, United States of America
[c] VVOB- education for development, Belgium
[d] UNU-WIDER, Finland

ARTICLE INFO

JEL classification:
I22
I28
O15

Keywords:
School leadership
Professional development
School management

ABSTRACT

Can effective school leadership enhance high-stakes test scores in low-income countries? To address this question, we examine the short-term impact of a school leadership professional development program that was implemented by VVOB across 525 primary schools in the six lowest-performing districts of Rwanda between 2018 and 2019. The program aimed to strengthen the leadership, management, and teacher support skills of school headteachers. We find that the program had small but statistically insignificant effects on Primary Leaving Examination scores within one to two years after the intervention. However, the program led to a five to six percentage point increase in teacher retention rates, with qualitative evidence suggesting that headteachers provided greater support to teachers, in particular during the onboarding process. Future research should focus on refining such programs and understanding which mechanisms are necessary to also improve learning outcomes of students.

1. Introduction

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).

In this regard, 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; Muralidharan and Sheth, 2016). 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 (Branch et al., 2012; Crawfurd, 2017; Muñoz and Prem, 2024). School leaders 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, three questions remain: Can high-quality training programs improve student learning outcomes? Are these programs cost-effective at scale in terms of learning gains (Anand et al., 2023; Bloom et al., 2015; Lemos et al., 2021)? Beyond student outcomes, can such programs also improve outcomes for teachers? 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 program 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. 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) and on teacher retentions rates. 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 small but statistically insignificant effects on student test scores. We can rule out average

* Corresponding author.
E-mail address: simeon.lauterbach@graduateinstitute.ch (S. Lauterbach).

https://doi.org/10.1016/j.jdeveco.2025.103545
Received 22 December 2024; Received in revised form 15 May 2025; Accepted 17 May 2025
Available online 17 June 2025

treatment effects on aggregated test scores of 0.09 standard deviations with 90% confidence and 0.10 standard deviations with 95% confidence. We also do not find significant and robust heterogeneous treatment effects, when it comes to the gender of students or ownership structure (that is, schools either operated by the government or private schools that receive government aid).

We also examine the effects of the SLPDP on teacher outcomes. It is plausible that the program led to broader changes within schools beyond student performance (Goldberg et al., 2019). Up to 35% of teachers leave their schools each year in Rwanda (Zeitlin, 2021). We assess whether the program improved retention, defined as the share of teachers who continue teaching at the same school the following year. We find a modest improvement of five to six percentage points in retention rates for treatment schools, which is statistically significant at the 10 percent level.

Qualitative interviews with headteachers and mathematics teachers, though not representative, suggest that the program improved headteachers' knowledge and confidence in school management. Teachers also reported receiving more support from headteachers after the program, particularly during the onboarding of new staff. While these findings point to a better working environment for teachers, they did not translate into higher student test scores.

We recommend further research into how SLPDPs influence school leadership practices and how these, in turn, shape the broader school environment, including teaching and learning conditions. Gaining such insights is essential for identifying the mechanisms through which SLPDPs can not only enhance working and teaching conditions but also translate into meaningful improvements in student learning outcomes. This is particularly important given that our calculations using the program's costs show that even modest learning gains, in the order of 0.05 standard deviations, an effect that we cannot rule given our estimates may be moderately cost-effective compared to alternative educational interventions (Anand et al., 2023; Kremer et al., 2013).

It is important to note that our evaluation analyses only the first cohort of the School Leader Professional Development Program. As the program evolves, refinements in its implementation may lead to better outcomes. It may also be possible that effects on student test scores take longer than one or two years to materialize, given the long causal chain from training, to school leader behavior change, to teacher behavior change (Anand et al., 2023; de Hoyos et al., 2020, 2021). These considerations suggest that early evaluations may underestimate the full potential impact of such programs and highlight the value of continued monitoring over time.

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 finds that school leadership and management programs have a small but insignificant impact of approximately 0.04 standard deviations on test scores in low- and middle-income countries (Anand et al., 2023). For example, de Hoyos et al. (2020) show that providing information on student ability to school leaders in Argentina increased test scores in mathematics and reading by 0.35 standard deviations. It also led to changes in curriculum, improved teacher quality assessment, and increased parental awareness about students performances. Also, a three-year management training program for school principals in Brazil increased student learning by 30% at minimal public costs (Barros et al., 2019). A more recent study in Malawi reports that test scores increased on the order of 0.10 standard deviations after a school management intervention. This was achieved mainly through the provision of remedial classes (Asim et al., 2024).

However, most studies do not record statistically significant learning gains. A randomized school management intervention in Madagascar shows 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 analyzing a large-scale training program, with a relatively high take-up rate among headteachers (around 70%) and a substantial number of test-takers (over 50 000 students). This allows us to measure precise treatment effects and explore more detailed heterogeneous treatment effects. Our findings are in line those of a recent meta-analysis of Anand et al. (2023) by ruling out large and significant effects on student-test scores. That review did though find effects of programs similar to the one that we study on school leader and teacher practices. The program we study is similar to other NGO-supported programs in covering a large number of research-based "high-quality practices". However, there are two important distinctions from similar training program. First, the program trained only one headteacher per school, whereas other programs, such as those in The Gambia or Malawi, also targeted additional actors, including teachers and sub-district officials. Second, while our program included a relatively high number of training days, it was delivered as a low-intensity intervention spread over 10 months. In contrast, the program in Malawi, for example, was implemented in a concentrated format over 10 consecutive days.

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 three studies, in Madagascar (Lassibille, 2016), in Malawi (Asim et al., 2024) and Gambia (Blimpo et al., 2015), have examined the effects of SLPDPs in Sub-Saharan Africa. We fill this gap by presenting for the first time evidence from a school leadership professional development program in Rwanda.

Moreover, we assess the impact of the School Leader Professional Development Program (SLPDP) using high-stakes test scores. Most studies rely on low-stakes assessments to measure learning outcomes, which, from the student's perspective, may not accurately reflect their true abilities. Research has shown that students often underperform on low-stakes assessments, not due to lack of ability but due to lower motivation (Finn, 2015). Gneezy et al. (2019) show that students from two high schools in the United States, without any incentives, perform worse than students from four high schools in Shanghai on low-stakes standardized tests. Once incentivized, US students significantly improve their performance, while Shanghai students, already top performers, show little change. This suggests that differences in test rankings may reflect not only ability but also varying levels of motivation to perform on low-stakes assessments. The studies on the effects of SLPDP in Malawi and Gambia use low-stakes exams (administered by the research teams) to evaluate the impact of their intervention. Both papers find small and significant effects on student test-scores. However, it is plausible to assume that headteachers who participated in the SLPDP were more motivated to encourage their teachers and students to perform well on the tests of the research teams, whereas headteachers who did not participate in the SLPDP lacked similar incentives. As a result, differences in test scores might stem from motivation issues rather than the SLPDP itself. Our study is the first to use test-score data from a high-stakes assessment, the Primary Leaving Examination in Rwanda, which determines students' transition to secondary education. In contrast to the aforementioned studies, we find that the SLPDP had no significant effect on student test scores.

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. Even though we do not find significant treatment effects, we contribute to the literature by conducting a thorough cost-effectiveness analysis, given that we have detailed data on the program's cost. Only four studies globally have reported exact program costs for school leadership professional development programs (Anand et al., 2023). Given that we can only rule out treatment effects in the order of 0.09 standard deviations or higher with 90% confidence, this analysis assumes that, if optimized accordingly, the SLDPD may achieve long-term learning gains of approximately 0.05 standard deviations. Our results indicate that, under this assumption, the SLDPD would rank within the medium effectiveness range of learning interventions.

2. Intervention

2.1. Education 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).

Overall, whilst Rwanda has made substantial progress in recent years in expanding participation and improving infrastructure, school quality remains a challenge. Enrollment rates surged following the 2008 and 2012 policies on nine and twelve years of free basic education. The primary gross enrollment 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). Yet while public and government-aided primary and secondary schools are free from tuition, parents still need to pay fees for mock exams, registration, and other costs, and these fees pose challenges for successful school attendance, performance, and completion (Trines, 2019). 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).

Of importance for our study is that there are three distinct types of ownership and management structures of schools: public, private, and government-aided schools. Out of 2.5 million children enrolled in 2909 primary schools, 32% were enrolled in public schools, 64% in government-aided private schools, and 4% in private schools (data from the 2018 Annual School Census). 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 typically founded and owned by faith-based organizations, predominately catholic or protestant churches. School management involves collaboration between the government and the faith-based organization (King, 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 the day-to-day operations of the school or teacher appointments (UNESCO, 2023). Government-aided schools often have better infrastructure, owned and funded by the church or NGO, and tend to perform better academically. Whilst Rwanda has long had a policy of free primary education, individual schools do typically ask for some contribution from parents for things such as materials and facilities. Parents of public primary school students spend around 10–15 percent more on school than parents of government-aided school students, with a similar gap in overall household income (NISR, 2024). Enrollment decisions for both public and government-aided schools are typically based primarily on proximity.

2.2. The school leader 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).[1] 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 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.[2] 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 CPD is an accredited diploma program focusing on "Effective School Leadership". It is delivered and accredited by the University of Rwanda – College of Education for Primary and Secondary Schools.

The educational curriculum is designed around the five professional standards for effective school leadership, encompassing vital aspects

[1] 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).

[2] 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.

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 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.

The theory of change follows a four-step pathway: (1) improved headteacher practices lead to (2) a better working environment for teachers, which in turn fosters (3) a more effective learning environment for students, which improves student learning outcomes, ultimately resulting in (4) improved student outcomes.

In this paper, we primarily test if the program had an effect on student test scores. In the mechanism section, we test quantitatively if the program improved teacher retention rates. We use teacher retention rates as a proxy for a supportive working environment for teachers (Burke and Sass, 2013; Cullen et al., 2021). Moreover, we analyze if improved teacher retention rates correlate positively and significantly with student test scores. This helps to understand whether teacher retention is positively associated with student performance. Moreover, we complement our quantitative findings with qualitative evidence from interviews and focus group discussions to better interpret the program's effects.

The program was set up for the first cohort 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 move on to another training centers and give the training there. Overall, there is minimal turnover among instructors ensuring that all headteachers experienced a similar level of training quality.

2.3. Randomization

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[3] in both primary and secondary schools in all 30 districts in Rwanda.[4] 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.[5] 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 the 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, but please note that none of the schools in cohort 0 were part of the randomization process for cohort 1,2 or 3.

The program aimed to provide training to headteachers in a total of 525 schools, spread across three cohorts.[6] Schools were allocated across the three cohorts by randomization, 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 type (primary, 9YBE, and 12 YBE). Out of the 591 primary schools, 66 had already participated in cohort 0, leaving 525 eligible schools. Each of these schools was 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, with the aim of having an equal number of schools per district. VVOB also ensured that during the randomization process they made sure to have a balanced number of schools across cohorts with regard to school-type and ownership, which Table A1 confirms. Fig. 1 displays the locations of the schools in cohort 1, 2, and 3 in the six chosen districts across Rwanda.

After having compiled a list of schools to be contacted for each cohort, headteachers were 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. In addition, 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

[3] 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 14 837 villages.

[4] 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.

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

[6] 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.

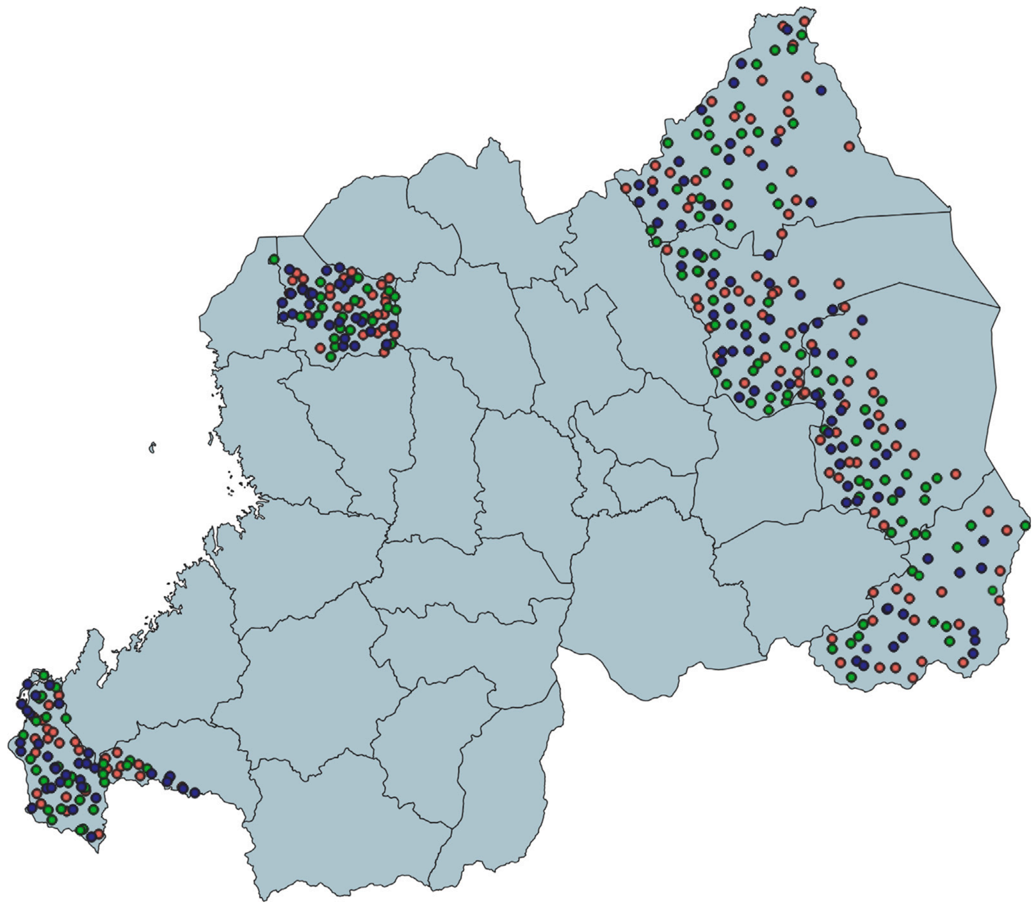

Fig. 1. Locations of primary schools in cohort 1 (red), cohort 2 (blue) and cohort 3 (green).
NOTE: This figure was computed by using the geolocations of schools that received or were supposed to participate in the SLPDP between 2018 and 2020.

from cohort 3 participated in cohort 2 or cohort 1. However, aside from the fact that complier schools tend to be slightly smaller, we do not observe any significant or meaningful differences in school characteristics between complier and non-complier schools (see Table A17). Thus, we estimate the intent-to-treat effects. 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 also received 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.

The timeline illustrated in Fig. 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.[7] 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.

Cohort 1 completed the training before the 2018 exams, while cohort 2 received partial training before the 2019 exams. To obtain a clean comparison 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.[8]

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

[8] 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 the intent-to-treat effect for test-scores, given that we have a type I error rate of 0.05, desired power of 0.80, inter-cluster correlation of 0.04 (estimated from the data), 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.

Fig. 2. Timeline of the program.

3. Identification

3.1. Data sources

We use various data sources for this paper. First, information on the selected schools (in cohort 1, 2, and 3), and information on participating school leaders were provided by VVOB. Data on the program participants include birth year, gender, education level, training dates and if they passed the final exam.[9]

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 obtained teacher placement data from the Ministry of Education for the years 2017, 2018, and 2019. This dataset covers all teachers on the government payroll in public and government-aided schools during the first term of each academic year (January–March). It includes details on each teacher's subject specialization, gender, and date of hire.

Fourth, we rely on several complementary datasets to assess school and sector characteristics. Our primary source for this purpose is the Annual School Census from 2017, which was collected during summer 2017 (MOE, 2017). This dataset includes key indicators such as access to electricity and water, the number of teachers and students, the number of students in P6, and the number of classrooms per school.

To enrich school-level data, we further incorporate information from the PRIO-GRID dataset, which provides sector-level indicators including nightlight intensity, gross cell product, and distance to Kigali. Last, we draw on the 2012 Population Census to add data on sector-level population density.[10]

We then merged the datasets to create a unified dataset for analysis. This process began with manually identifying schools in the test score dataset that were part of the randomization, using school, district and sector names. This procedure is relatively accurate given that there are on average six to eight schools per sector and school names within sectors are usually quite different. Once we identified the intervention schools in the test-score data, we merged the relevant intervention information from VVOB into the test score data.

We were able to locate nearly all government-aided and public schools in the test-score dataset. For instance, out of the 274 government-aided schools that were randomly assigned to the treatment or control group, we successfully matched 257 schools (94%) 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 155 schools (91%) with the test-score data.[11]

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.

Given the low matching rate for private schools and the fact that only about 4% of students in Rwanda are enrolled in them, we opted to exclude private schools from our main analysis. As a result, our primary focus is on public and government-aided schools.[12]

This decision is unlikely to bias our estimates, as the randomization process ensured a balanced number of private schools across cohorts. Table A18 shows that private schools are significantly smaller, newer, and have better access to electricity and water compared to public and government-aided schools. Therefore, we interpret the results presented in this paper as applicable to medium- to low-resource schools. Nevertheless, to assess the robustness of our results, we perform a sensitivity analysis that includes the few matched private schools to examine whether the program may have had differential effects on this subgroup, based on the limited data available.

We then merged the annual school census data,[13] the Prio Grid data, as well as the population census to our test-score data using either school, sector and district names or school codes. We verified manually that the merging procedure was correct.

Last, we construct a teacher retention rate index for each public and government-aided that could be successfully matched with the test-score data. We do this for the years 2017 to 2018 and 2018 to 2019. The teacher retention rate captures the share of teachers who remained at the same school from one year to the next.

Since the teacher data for each year (2017, 2018, and 2019) was collected during the first term of the academic year (January-March), we are able to compute a pre-intervention teacher retention rate 2017–2018 and a post-treatment teacher retention rate 2018–2019. The latter therefore captures how many teachers stayed in the same school up to five month after the SLPDP ended.[14] To compute the teacher retention rate, we did the following:

First, we manually identified all schools in the teacher dataset that participated in the intervention and are also present in the test score dataset. Of the 412 public and government-aided intervention schools included in the test score data, 401 schools (97%) appear in the teacher

[9] Please note that over 80% of respondents passed the final exam.

[10] Population density reflects the average number of inhabitants per square kilometer in a given area. The data we use comes from Rwanda's fourth Population and Housing Census conducted in 2012. While more recent figures were collected in 2022, they fall after the intervention period and are therefore not included in our analysis.

[11] There are several reasons why we were unable to match all schools to the test score data. In some cases, the spelling of school names differed between the VVOB dataset and the test score dataset. In other instances, test score data may not have been collected for certain schools or may simply be missing from our test-score dataset.

[12] In addition, attrition among public and government-aided schools is similar across cohorts. Combined with random assignment and baseline controls, this suggests that the exclusion of unmatched private schools should not introduce bias.

[13] For some schools in the 2017 Annual School Census, certain information, such as the number of P6 classrooms is missing. For these schools we impute the missing values using sample averages and include a dummy variable in the regression indicating whether the information is missing.

[14] We acknowledge that any effects on teacher retention may not be fully observable by March 2019. Unfortunately, due to COVID-19, no teacher data was collected in 2020, preventing us from assessing medium-term impacts of the intervention.

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%)	136/145 (94%)	140/149 (95%)
Compliance with the program:			
Received training in 2018	99/136 (73%)	7/136 (5%)	10/140 (7%)
Received training in 2019	7/136 (5%)	40/136 (29%)	33/140 (24%)
Received training in 2020	0/136 (0%)	0/132 (0%)	0/140 (0%)

dataset in 2017, 391 schools (95%) in 2018, and 392 schools (96%) in 2019.

Second, our teacher data from 2017, 2018 and 2019 does not have teacher names or IDs due to confidentiality. However, it does contain for each school the hiring data of the teacher and the gender. To track individual teachers across years, we define a teacher as uniquely identifiable if no other teacher of the same gender was hired on the same date at the same school. Using this combination of school name, gender, and hire date, we identify and match teachers across the 2017, 2018, and 2019 datasets.

Third, we calculate teacher retention rates by assessing how many uniquely identified teachers in 2017 were still present in the 2018 teacher data, and how many uniquely identified teachers in 2018 remained in 2019.[15] Also for the few intervention schools in the test score data that we could not merge to the teacher data, we impute their teacher retention rate between 2017 and 2018 and 2018 and 2019 with the mean for those years.

Overall, we find an average teacher retention rate of 66% between 2017 and 2018, and 68% between 2018 and 2019. These figures closely align with the estimate reported by Zeitlin (2021), who calculates a retention rate of approximately 70% based on data from 250 public primary schools in Rwanda, thereby lending credibility to our results. We further discuss the interpretation and implications of this retention rate in the mechanism section of the paper.

The final data-set appears as following: Out of the 149 public and government aided schools from cohort 1 schools, 136 were successfully merged to the test-score data, which corresponds to a matching rate of 91%. Similarly, in cohort 2, 136 out of the total 145 public and government aided schools were merged (matching rate of 94%) and in cohort 3, 140 out of the total 149 public and government aided schools were merged (matching rate of 95%).[16]

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.

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:

$$Y_{istd}^{j} = \alpha + \beta \times Treatment_s + X_{istd} + \sum_{z=2015}^{2017} \omega_z \times \bar{Y}_{s,z}^{j} + Z_s + \lambda_d + \Omega_t + \epsilon_{istd}, \quad (1)$$

where the outcome of interest, denoted as Y_{istd}^{j}, represents the individual student's performance i in subject j on the primary leaving examination in year t (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.[17] We standardize the grades in the treatment and control group with mean 0 and a standard deviation of 1.[18]

$Treatment_s$ 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_{istd} 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). We also include the population density of the sector the school is located in, given that there is a slight imbalance between the treated and control cohort. We follow McKenzie (2012) and include the average grade in the subject of interest per school, denoted as $\bar{Y}_{s,z}^{j}$ one, two and three years before the intervention took place. This approach helps to increase power and precision.[19] We also include district fixed effects λ_d, test year fixed effects ω_t and cluster standard errors at the district level.

In summary, the standard regression includes controls for students' gender and birth year, the population density of the school sector, school ownership, school type, test year and district fixed effects, as well as the average grades in the relevant subject one, two, and three years prior to the intervention.

To increase power and precision we run another specification that includes, next to the standard regression model, additional control variables at the school level Z_s. We employ the standard lasso approach, using cross-validation (CV) to select the optimal regularization parameter (λ). To ensure reproducibility, we set the random seed to 12 345. Through lasso, we identify the following control variables as highly relevant: the teacher retention rate between 2017 and 2018, the number of students in P6, the number of students taking the PLE exam in 2018 and 2016, the number of students in P6, whether a school was opened after the Genocide against the Tutsi in 1994, the availability of electricity or water, whether a school has a school feeding program, the distance to the capital and the gross cell product in the sector of the school.[20] Therefore, whenever we refer to additional controls in the table, we include the above mentioned set of control variables that we found through Lasso.

For the heterogeneity analysis, we add an interaction term between $Treatment_s$ and the gender of the student or whether the school is a public school. In the heterogeneity analysis section, we explain why we selected these two variables. We address the issue of multiple hypothesis testing by adjusting the p-values using the Bonferroni correction method.

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

[15] For example, if a school had 10 uniquely identified teachers in 2017 and 8 of them could still be identified in the 2018 dataset, we compute a retention rate of 0.8 for that year.

[16] We also conduct a balancing test on schools from cohort 1 and cohort 3 that we were unable to merge to the test-score. The results of this test, shown in Table A2 in Appendix A, indicate that the characteristics of the unmatched schools are relatively similar, giving us confidence that we are not systematically excluding specific types of schools from our analysis.

[17] 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.

[18] 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.

[19] We do not have the grade information for 4 schools in 2017, 2016 and 2015. We replace this missing information with the average grade in subject j from 2017, 2016 or 2015.

[20] 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).

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 assume that the non-interference, excludability and monotonicity assumptions are satisfied.[21] In a first step, the random allocation across cohorts is used as an instrument for actual participation in the school leadership training program:

$$P_s = \alpha + \beta \times Treatment_s + X_{istd} + \sum_{z=2015}^{2017} \omega_z \times \bar{Y}^j_{s,z} + Z_s + \lambda_d + \Omega_t + \epsilon_{istd}, \quad (2)$$

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:

$$Y^j_{istd} = \alpha + \beta \times \tilde{P}_s + X_{istd} + \sum_{z=2015}^{2017} \omega_z \times \bar{Y}^j_{s,z} + Z_s + \lambda_d + \Omega_t + \epsilon_{istd}, \quad (3)$$

4. Results

4.1. Balance tests and summary statistics

This section presents summary statistics and assesses balance between schools in cohort 1 (treatment) and cohort 3 (control) in our final school sample. For the balancing, we use test score data from 2017, 2016 and 2015, so before the intervention took place, the Annual School Census from 2017 and various sector characteristics that were derived from the PRIO-GRID data and the population census.

Schools in cohort 1 (treatment group) and cohort (3) are well balanced, when it comes to school and student characteristics. Looking at school characteristics in Table 2 shows that there is no imbalance with respect to nearly all school characteristics between the treatment and control group.

Schools in the control group are situated in slightly more densely populated areas and more students took the PLE exam in 2017.[22] Overall, the F-test of joint significance is insignificant with an F-value of 1.21 and a p-value of 0.23.

Looking at the characteristics of students in treatment and control schools in the years before the intervention started, we see no statistically significant differences in test-scores, gender, and age of taking the PLE exam in 2017, 2016 and 2015 (Table 3).

4.2. Average treatment effects

This section reports the average intent-to-treat effects, before discussing heterogeneity. Note that the SLPDP did not significantly change the composition of students attending 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 on pooled test-scores from 2018 and 2019.

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. When looking at the first columns, we can rule out average treatment effects of 0.09 standard deviations with 90% confidence and 0.10 standard deviations with 95% confidence. We can also rule out potential spillover effects by examining whether test scores increased in cohort 3 schools located near cohort 1 schools or in cohort 1 and 3 schools situated close to pilot phase (cohort 0) schools. In theory, spillover effects would bias our main estimates in Table 4 downward. For the spillover effects please refer to section A.5. As the coefficients in columns (1) to (8) are estimated with small standard errors, we conclude that the SLPDP intervention has a precise zero effect for test-scores in 2018 and 2019.

We further test whether there are different treatment effects in 2018 or 2019 that may not be captured by the pooled data in Table 4. For example, treatment effects might only become evident a year after the intervention, in our case, in 2019. Such delays could be due to various factors, such as headteachers' inability to implement new structures or recruit additional teachers until the following academic year.

Table A4 presents the treatment effects for 2018, while Table A5 shows the effects for 2019. None of the coefficients in either table are statistically significant. However, the point estimates for 2019 are generally higher than those for 2018. For instance, the average treatment effect on aggregate test scores is 0.03 to 0.04 standard deviations in 2019 (columns 1 and 2 in Table A5) and 0.00 to 0.01 in 2018 (columns 1 and 2 in Table A4). This could be an indication that the program's effects only appear later, but the statistical significance is too small to draw any meaningful conclusions. Last, as a sensitivity analysis, we examine whether the average treatment effects change when including the private schools that were initially excluded from the analysis. As shown in Table A6, all coefficients remain stable and unchanged.

In summary, this section finds no meaningful effect of the intervention on test scores, either in the short or long term, when estimating average treatment effects. The coefficient sizes, particularly in Table 4, are comparable to the average (insignificant) effect size of 0.04 standard deviations reported in the meta-analysis of school leader training programs by Anand et al. (2023). However, we cannot rule out the SLPDP could have longer-term effects. In addition, as some schools in cohort 3 received partial training in 2019, this could introduce a downward bias in the observed treatment effects.

4.3. Heterogeneous effects

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

In Rwanda there is gender parity in primary-level enrollment, but 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 tend to achieve lower learning outcomes. The exact mechanisms are not fully understood, but female students may be at a disadvantage due to more family responsibilities, or a lack of sanitary pads during menstruation (MOE, 2018; Nzaramba et al., 2021; UNICEF, 2024). Our focus revolves around investigating

[21] The non-interference 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.

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 - see Appendix A.5.

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.

[22] We decided to include population density as a control variable in the main regression, given its theoretical relevance. Since there is no evidence of imbalance in the number of test takers across the years 2016, 2018, and 2019, we allow the LASSO procedure to select which of the available test-taker counts from 2016, 2017, 2018, or 2019 should be included as additional controls.

Table 2
Baseline descriptive and balance.

Variable	Sample mean	Treatment	Control	Mean diff.	P-value	Size treatment	Size control
School characteristics mainly from the Annual School Census in 2017							
Government aided	0.61 (0.49)	0.60 (0.49)	0.62 (0.49)	−0.02	0.67	136	140
Primary	0.61 (0.49)	0.60 (0.49)	0.62 (0.49)	−0.02	0.77	136	140
9 YBE	0.23 (0.42)	0.25 (0.43)	0.21 (0.41)	0.03	0.52	136	140
12 YBE	0.16 (0.37)	0.15 (0.36)	0.16 (0.37)	−0.02	0.72	136	140
Total students in 2017	1227.23 (768.59)	1173.36 (706.32)	1279.57 (823.76)	−112.96	0.19	136	140
Students in P6 in 2017	111.62 (68.19)	106.90 (66.76)	116.21 (69.49)	−9.69	0.20	136	140
PLE takers in 2019	110.28 (66.42)	105.04 (66.52)	115.38 (66.15)	−11.67	0.11	136	140
PLE takers in 2018	97.36 (63.83)	93.05 (63.48)	101.54 (64.12)	−9.07	0.20	136	140
PLE takers in 2017	89.29 (57.40)	81.74 (49.30)	96.64 (63.62)	−14.78	0.02**	136	140
PLE takers in 2016	73.89 (48.22)	69.63 (43.21)	78.02 (52.46)	−8.04	0.15	136	140
P6 classrooms in 2017	2.69 (2.19)	2.79 (2.62)	2.60 (1.68)	0.19	0.48	136	140
Teachers in 2017	18.92 (13.06)	18.18 (12.47)	19.65 (13.61)	−1.56	0.29	136	140
Teacher retention 2017–2018	0.64 (0.34)	0.65 (0.34)	0.62 (0.34)	0.03	0.375	136	140
Establishment date	1981 (24.57)	1982 (24.64)	1980 (24.56)	1.41	0.63	136	140
Establishment after 1994	0.43 (0.50)	0.46 (0.50)	0.41 (0.49)	0.05	0.35	136	140
Electricity 2017	0.51 (0.50)	0.49 (0.50)	0.53 (0.50)	−0.04	0.48	136	140
Water 2017	0.55 (0.50)	0.52 (0.50)	0.58 (0.50)	−0.05	0.38	136	140
School feeding program	0.30 (0.46)	0.32 (0.47)	0.29 (0.45)	0.03	0.54	136	140
Distance to the capital in km	233.47 (464)	228.60 (456)	239.18 (473)	−0.62	0.99	136	140
Gross Cell product in 2005	0.12 (0.06)	0.12 (0.06)	0.12 (0.06)	0.00	0.96	136	140
Night light activity in 2012	0.96 (1.36)	0.89 (1.33)	1.03 (1.38)	−0.03	0.77	136	140
Night light activity in 2012 above median	0.51 (0.50)	0.50 (0.50)	0.53 (0.50)	0.01	0.83	136	140
Urban	0.07 (0.26)	0.07 (0.26)	0.07 (0.26)	0.00	0.81	136	140
Pop density 2012	463.13 (463.52)	429.07 (429.33)	496.22 (496.22)	−59.31	0.03**	136	140
Pop density above median	0.50 (0.50)	0.45 (0.50)	0.55 (0.50)	−0.08	0.11	136	140
F-test of joint significance	F-value: 1.21	P-value: 0.23					

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 A3. The main data sets for this table are coming from the Annual School Census data. The variables night light activity and population density are coming from the PRIO-GRID data sets and the number of students taking the PLE exit exam from the test-score data sets.

the potential impact of the SLPDP in enhancing female test scores relative to male test scores.

We also hypothesize that pre-existing management practices may be different between public and government-aided schools. Evidence from other countries is mixed, with the World Management Survey showing that "autonomous government schools" tend to exhibit superior management practices in several countries (Bloom et al., 2015), but in neighboring Uganda there are minimal differences in average management quality between public and government-aided schools (Crawfurd, 2017). As discussed in Section 2, the funding and governance structures of public and government-aided schools differ significantly. Given these structural differences, our study aims to investigate whether the School Leadership and Professional Development Program (SLPDP) has differential effects on these two types of schools. It is plausible that the impact of such interventions may vary depending on the operational structure of the school.

We perform the heterogeneity analysis on test scores pooled across 2018 and 2019 to increase power and precision. As before, we conduct these regressions on aggregated test scores in columns (1)–(2), English test scores in columns (3–4), non-stem test scores in columns (5–6) and stem test scores in columns (7)–(8).

Table 5 displays the impact of the SLPDP on male students, as denoted by "Treatment", while "Treatment × Female" depicts how the treatment effect differs between male and female students. First, the table confirms previous research highlighting the significant and persistent gender gap in learning outcomes between female and male students. More specifically, male students outperform female students by 0.20 to 0.25 standard deviations. Second, the SLPDP did not succeed

Table 3
Baseline descriptive and balance (continued).

Variable	Sample mean	Treatment	Control	Mean diff.	P-value	Size treatment	Size control
Student test scores from the test score data in 2017							
Aggregate grade 2017	0.00	0.06	−0.05	0.09	0.13	11 117	13 440
	(1.00)	(1.02)	(0.98)				
Stem grade 2017	0.00	0.04	−0.03	0.06	0.28	11 117	13 440
	(1.00)	(1.01)	(0.99)				
Non-Stem grade 2017	0.00	0.05	−0.04	0.08	0.22	11 117	13 440
	(1.00)	(1.02)	(0.98)				
English grade 2017	0.00	0.04	−0.03	0.06	0.29	11 117	13 440
	(1.00)	(1.02)	(0.98)				
Birth year	2003	2003	2003	−0.02	0.67	11 117	13 440
	(1.70)	(1.72)	(1.68)				
Female student	0.55	0.54	0.55	−0.01	0.36	11 117	13 440
	(0.50)	(0.50)	(0.50)				
F-test of joint significance	F-value: 1.27	P-value: 0.27					
Student test scores from the test score data in 2016							
Aggregate grade 2016	0.00	0.01	−0.01	0.01	0.84	9397	10 923
	(1.00)	(0.98)	(1.01)				
Stem grade 2016	0.00	0.00	−0.00	−0.00	0.96	9397	10 923
	(1.00)	(0.99)	(1.01)				
Non-Stem grade 2016	0.00	0.01	−0.01	0.01	0.84	9397	10 923
	(1.00)	(1.00)	(1.00)				
English grade 2016	0.00	0.01	−0.01	0.02	0.78	9397	10 923
	(1.00)	(0.98)	(1.01)				
Birth year	2002	2002	2002	−0.04	0.42	9397	10 923
	(1.64)	(1.63)	(1.65)				
Female student	0.55	0.55	0.55	−0.00	0.87	9397	10 923
	(0.50)	(0.50)	(0.49)				
F-test of joint significance	F-value: 0.34	P-value: 0.92					
Student test scores from the test score data in 2015							
Aggregate grade 2015	0.00	0.03	−0.03	0.05	0.36	8183	9434
	(1.00)	(0.99)	(1.01)				
Stem grade 2015	0.00	0.01	−0.01	0.02	0.79	8183	9434
	(1.00)	(0.98)	(1.02)				
Non-Stem grade 2015	0.00	0.04	−0.04	0.07	0.27	8183	9434
	(1.00)	(0.99)	(1.00)				
English grade 2015	0.00	0.02	−0.02	0.04	0.55	8183	9434
	(1.00)	(0.98)	(1.01)				
Birth year	2001	2001	2001	0.00	0.65	8183	9434
	(1.65)	(1.63)	(1.66)				
Female student	0.54	0.54	0.54	0.01	0.59	8183	9434
	(0.59)	(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, we have test-score information for 275 out of the 276 schools in 2017 and 2016 and for 271 out of the 276 schools in 2015.

Table 4
Treatment effects on test-scores in 2018 and 2019.

	(1) Aggregate	(2) Aggregate	(3) English	(4) English	(5) Nonstem	(6) Nonstem	(7) Stem	(8) Stem
Treatment	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Add controls	No	Yes	No	Yes	No	Yes	No	Yes
Number of students	57,309	57,309	57,309	57,309	57,309	57,309	57,309	57,309
Number of schools	276	276	276	276	276	276	276	276
R^2	0.19	0.20	0.16	0.17	0.19	0.19	0.16	0.17

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 test year fixed effects, district fixed effects, the birth year, 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 2017, 2016 and 2015. Additional controls include the teacher retention rate between 2017 and 2018, the number of students in P6, the number of students taking the PLE exam in 2018 and 2016, the number of students in P6, whether a school was opened after the Genocide against the Tutsi in 1994, the availability of electricity or water, whether a school has a school feeding program, the distance to the capital and the gross cell product in the sector of the school. 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.

Table 5
Treatment effects for female students on test-scores in 2018 and 2019.

	(1) Aggregate	(2) Aggregate	(3) English	(4) English	(5) Nonstem	(6) Nonstem	(7) Stem	(8) Stem
Treatment	0.04	0.04	0.04	0.05	0.04	0.04	0.06	0.05*
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
Female	-0.20^{***}_{***}	-0.20^{***}_{***}	-0.11^{***}_{***}	-0.12^{***}_{***}	-0.22^{***}_{***}	-0.22^{***}_{***}	-0.25^{***}_{***}	-0.25^{***}_{***}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment × Female	−0.02	−0.03	−0.02	−0.02	−0.02	−0.02	−0.05*	−0.05*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Add controls	No	Yes	No	Yes	No	Yes	No	Yes
Number of students	57,309	57,309	57,309	57,309	57,309	57,309	57,309	57,309
Number of schools	276	276	276	276	276	276	276	276
R^2	0.19	0.20	0.16	0.17	0.19	0.20	0.16	0.17

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. For more information about the control variables used we refer to Table 4. Reported are unadjusted p-values* and adjusted p-values$_{*}$ for multiple hypothesis testing. All outcome variables in the table are grouped as one family.

Table 6
Treatment effects for public schools on test-scores in 2018 and 2019.

	(1) Aggregate	(2) Aggregate	(3) English	(4) English	(5) Nonstem	(6) Nonstem	(7) Stem	(8) Stem
Treatment	−0.03	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02
	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)	(0.4)	(0.04)	(0.03)
Public	−0.03	−0.06	−0.05	−0.09*	−0.02	−0.05	−0.04	−0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Treatment × Public	0.15**	0.13*	0.14*	0.13*	0.13*	0.11	0.13**	0.12*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.06)	(0.06)
Add controls	No	Yes	No	Yes	No	Yes	No	Yes
Number of students	57,309	57,309	57,309	57,309	57,309	57,309	57,309	57,309
Number of schools	276	276	276	276	276	276	276	276
R^2	0.19	0.20	0.16	0.17	0.19	0.19	0.16	0.17

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. For more information about the control variables used we refer to Table 4. Reported are unadjusted p-values* and adjusted p-values$_{*}$ for multiple hypothesis testing. All outcome variables in the table are grouped as one family.

in reducing this gap. The treatment effect for girls is 0.02 to 0.05 standard deviations smaller than the effect for boys. This difference is not statistically significant when adjusting for multiple hypothesis testing.

Table 6 indicates that the SLPDP had a stronger impact on public schools compared to government-aided schools. The point estimates in columns (1) through (8) are slightly negative for public schools, though they are not statistically significant.

In contrast, the difference in the main effect between public and government-aided schools is significant across all columns, except for column (6), with the effect size ranging from 0.12 to 0.15 standard deviations, significant at the ten and five percent levels. However, the significance levels observed for public schools disappear when adjustments are made for multiple hypothesis testing. As a result, there is no significant and robust evidence that our SLPDP worked better in public schools. Further research is needed to determine whether our estimated coefficients are simply the result of chance or if the SLPDP genuinely performed better in public schools.

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 on both test-scores in 2018 and 2019.

Table 7 reports the results of the LATE estimates. The robust F-statistic is well above the common threshold of ten, suggesting that our instrument is significant. The LATE estimates are similar in magnitude to the ITT effects and insignificant. This suggests again that the overall impact of the school leadership professional development program is insignificant. Tables A8 and A9 show the LATE estimates separately for the test-scores in 2018 or 2019. There is no evidence that the program may have worked differently on test-scores in 2018 or 2019.

4.5. Placebo tests and alternative explanations

4.5.1. Placebo average treatment effects

We proceed by computing placebo average treatment effects to rule out that the sample of students for treated and control schools could have changed significantly in 2018 and 2019 after incorporating the control variables used in the previous specifications. Table A7 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. Student enrollment

Another concern is the possibility that the program could have influenced weaker students to remain enrolled in P6 and to 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. To address this issue, we regress the number of students taking the PLE exam in 2018 or 2019 on the treatment group. Table A10 shows that the number of students taking the PLE exam in treated schools may have increased by 4.54 in 2018 and by 1.00 in 2019. 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 4.54 students would represent a 3.82% percentage point rise in the number of students

Table 7
Local average treatment effects on test-scores in 2018 and 2019.

Variable	Aggregate	Aggregate	English	English	Nonstem	Nonstem	Stem	Stem
Treated (LATE)	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)
First-stage								
Treatment assignment	0.66***	0.65***	0.67***	0.66***	0.66***	0.65***	0.66***	0.65***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
H_0: weak instrument								
Robust F-statistic	165.25	181.66	180.55	198.71	171.15	187.55	165.10	185.36
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 and 2019	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.03
Number of students	57,309	57,309	57,309	57,309	57,309	57,309	57,309	57,309
Number of schools	276	276	276	276	276	276	276	276
R^2	0.19	0.20	0.16	0.17	0.19	0.19	0.16	0.17

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. For more information about the control variables used we refer to Table 4. Standard errors are clustered at the school level. ITT records the intent-to-treat effect as estimated previously for 2018 and 2019.

taking the PLE exam. This estimate is relatively modest and would, therefore, not significantly bias our estimates downwards.

We also conduct an additional robustness check by calculating Lee bounds for test-scores in 2018 and 2019.[23] As above, we assume that weaker students remain in schools implying a downward bias on test scores. Drawing upon our earlier estimations in table A10, we hypothesize that the SLPDP induced 5 students to take the P7 exam in treated schools in 2018 and for simplicity also in 2019. These students would have dropped out without the SLPDP.

Tables A11 and A12 present the lower and upper bounds of the treatment effects in 2018 or 2019, along with their 95% confidence intervals, when dropping the five highest and lowest performing students in treatment schools.[24] 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 exist. Therefore, we can confidently state that our previous regressions do not underestimate the actual treatment effect (Tauchmann, 2014).

5. Mechanisms

So far, our analysis has primarily focused on the impact of the School Leadership Professional Development Program (SLPDP) on student test scores. In this section, we aim to explore whether the program also impacted headteachers' practices and if the program had an impact on the teaching and learning environment. Answering these questions helps to identify potential breakdowns in the theory of change. Did the program improve the working environment for teachers, but this improvement did not translate into better test-scores or did the program fail to improve the working environment for teachers and therefore did not have an impact on student test-scores?

We acknowledge that we can only partially address these questions. Our analysis of changes in headteacher practices relies exclusively on qualitative interview data. In contrast, our assessment of changes in the teacher working environment draws on both qualitative interviews and quantitative evidence, providing a more robust empirical basis for this outcome.

5.1. The effects of the SLPDP on headteachers

We begin by examining how the SLPDP impacted headteachers' practices. To assess this, VVOB conducted in-depth interviews with seven headteachers participating in the SLPDP in 2018.

Headteachers reported that the SLPDP primarily enhanced their understanding of school management and leadership strategies. This new knowledge boosted their confidence and satisfaction, enabling a shift toward more collaborative decision-making. One headteacher shared: *"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"*.

The program also fostered a more hands-on approach to school management. Headteachers increasingly participated in teaching activities, providing essential support and feedback to teachers. For example, several headteachers emphasized the importance of structured teacher onboarding processes and improved curriculum implementation. One headteacher highlighted this transformation: *"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"*.

In addition, headteachers noted improvements in stakeholder engagement, particularly with parents and community members. Six out of seven of the interviewed headteachers reported regular parent meetings, which marked a shift toward a more inclusive school environment. This proactive engagement with stakeholders was accompanied by an increase in team spirit among school leaders, teachers, and students.

The program also encouraged gender-responsive practices in schools. Most of the headteachers reported that they focused more on reducing female student dropout rates and promoting gender equality in classrooms. In particular, several headteachers highlighted improvements in teaching strategies that supported female students' learning outcomes.

Despite these positive changes, headteachers faced significant challenges in implementing the SLPDP effectively, with time constraints being the most frequently cited issue. The multiplicity of responsibilities, ranging from administrative tasks to liaising with stakeholders, left headteachers struggling to dedicate sufficient time to teacher support. As one headteacher explained: *"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"*.

These challenges underline the need for more comprehensive support structures. For instance, headteachers suggested that organizations implementing the SLPDP should train additional school actors or engage other stakeholders, such as school owners, to ease the burden on headteachers and to facilitate the program's objectives.

[23] We refer to Tauchmann (2014) for further information.

[24] We run the Lee bounds separately for test scores in 2018 and 2019, as it is not possible to include test-year fixed effects in these calculations. In addition, we also exclude the five highest-scoring students to avoid placing any assumptions on the direction or size of potential bias.

Table 8
Treatment effects on the teacher retention between 2018 and 2019.

Variable	(1) Retention 2018–2019	(2) Retention 2018–2019	(3) Retention 2018–2019	(4) Retention 2018–2019
Treatment	0.05*	0.06*	0.05*	0.06*
	(0.03)	(0.03)	(0.03)	(0.03)
Mean	0.66	0.66	0.66	0.66
Cluster level	Robust	Robust	Sector	Sector
Add. controls	No	Yes	No	Yes
Number of schools	276	276	276	276
R^2	0.33	0.35	0.33	0.35

Note: Levels of significance: * < 0.10; ** < 0.05 *** < 0.01. We report treatment effects along with their associated standard errors. The teacher retention rates computes the proportion of teachers who began teaching at a school during the first term of 2018 and were still employed at the same school by the first term of 2019. For 18 schools, this retention data was unavailable and thus imputed using the sample average. Our standard regression model includes district fixed effects, school type, and ownership, as well as the density of the school's sector. The model also accounts for the teacher retention rate from 2017 to 2018 and includes a dummy variable to indicate whether the retention data for 2018 to 2019 was imputed. Additional controls in our analysis include the number of students in Grade 6 (P6), the number of students taking the Primary Leaving Examination (PLE) in 2018 or 2016, the number of classrooms in P6, and several school infrastructure variables: presence of water, electricity, a school feeding program, and the distance to the capital. We also control for the Gross Cell Product of the school's sector, the average aggregate grade of students in 2017, 2016, and 2015, and a dummy variable indicating if any school data is missing from the Annual School Census 2017. Standard errors are calculated to be either robust or clustered at the sector level.

5.2. The effects of the SLPDP on teachers

We now examine the program's effects on teachers, using both quantitative data and qualitative insights. We begin by examining the impact of the SLPDP on teacher retention rates from 2018 to 2019. This rate indicates the proportion of teachers who were employed at the beginning of 2018, coinciding with the start of the intervention, and who remained in their positions by early 2019, approximately five months after the intervention. As previously mentioned, we hypothesize that a higher teacher retention rate can be associated with an improved working environment for teachers.

Zeitlin (2021) estimates that approximately 20% of primary school teachers in Rwanda do not continue at the same school the following year.[25] He notes that about half of teachers departures are due to transfers to other schools, while the remaining half leave the teaching profession entirely. Consequently, we aim to test whether retention rates in schools that underwent the intervention improve compared to those that did not. It is important to note that we lack specific information on the reasons teachers leave their schools.

We apply Eq. (1) at the school level, focusing on the teacher retention rate from 2018 to 2019 as our dependent variable. We use a similar set of control variables as those used in previous analyses. Important, we control for the teacher retention rate prior to the intervention (from 2017 to 2018). We employ robust standard errors or cluster standard errors at the sector level.

Table 8 indicates that the teacher retention rate increased by five to six percentage points significant at the ten percent level. Considering a baseline retention rate of 66%, this represents a modest but notable improvement. We further corroborated our findings by running the same regression on the teacher retention from 2017 to 2018, so prior to the intervention, and find no effect as shown in Table A15. This results supports the hypothesis that the working environment may have improved in schools that were part of the intervention. Table A16, however, shows that higher teacher retention rates between 2017 and 2018 or between 2018 and 2019 do not exhibit a positive and statistically significant correlation with student test scores in 2018 or 2019. This suggests that teacher retention alone may not be sufficient to improve learning outcomes. Further research is needed to identify the conditions under which improvements in the teacher working environment translate into better student achievement.

[25] Please note that his rate rises to between 30 and 40% in the districts we are focusing on, as detailed in Fig. 2 of his paper.

Also, the interviews conducted with the teachers corroborate the findings above. During the interviews, teachers revealed several positive changes that occurred within their school after the headteacher participated in the SLPDP. 12 out of 30 teachers reported that their headteachers now facilitated more structured introductions for new teachers. This includes meetings with colleagues and giving more guidance with regard to classroom management. Previously, new teachers were often sent to classrooms with minimal preparations. Moreover, many teachers recognized the headteachers' efforts to improve access to teaching materials and to expand professional development opportunities. Both aspects may contribute to improve teaching quality.

This, however, did not lead to improved learning outcomes for students. A significant proportion of teachers (18 out of 30) still mentioned the lack of mentorship and coaching from headteachers during the school year. While induction programs were implemented, ongoing support beyond these programs remained insufficient. Teachers emphasized the need for additional time and professional development opportunities to implement the curriculum effectively. Teachers also highlighted resource constraints, including limited access to teaching materials such as math textbooks. As one teacher noted: *"The headmaster can only provide teaching materials upon availability, and overall support is still insufficient"*.

Regarding student learning, 26 out of 30 teachers reported noticeable improvements, particularly in mathematics, with significant progress observed among female students. We, however, cannot support this finding with our data at hand.

To summarize, the findings on the teacher retention rate, as well as the interviews conducted with headteacher and teachers suggest that the SLPDP had a positive effect on the teaching and working environment for teachers. This, however, did not translate into improved learning outcomes, as other systemic challenges, such as time constraints, resource shortages, and insufficient ongoing support for teachers remained. These challenges may explain the limited effects observed on student test-scores.

6. Cost-effectiveness analysis

Given that VVOB provided us with the exact costs of the program, we are still interested in conducting a cost-effectiveness analysis to determine whether programs like this can be scaled up in a cost-efficient manner. Although we do not find significant average treatment effects from the SLPDP, it is still possible that future cohorts may have experienced test score gains due to improvements within the program. In addition, we cannot rule average treatment in the order of less than 0.09 standard deviations with 90% confidence. To compare the costs and benefits of the program, we follow the methodology proposed by Bhula et al. (2023), JPAL (2023) and Walter (2020).

Based our insignificant test-score results, we simply assume for now that the school leadership professional training program has the potential to increase test-scores by 0.05 standard deviations. This estimate is also in line with the meta analysis conducted by Anand et al. (2023).

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. 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.[26]

As previously mentioned, we only have PLE test-score data and are therefore unable to estimate 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 or government aided school may have experienced an improvement of approximately 0.05 standard deviations.[27]

The average cost of the program was USD 1412 in 2018 per participating headteacher comprising substantial and administrative costs.[28] 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.

In a typical school of 850 students, cumulative learning gains would be $850 \times 0.05 = 42.5$ standard deviations. Translating these costs and benefits into total standard deviations of learning gains per USD 100, results into learning gains of 3.36 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.[29]

Assuming that the program led to an increase in test-scores by 0.05 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 0.0625 per student. By multiplying 0.0625 with the total number of students in a primary school (850) and dividing the result by 36.37, we arrive at a LAYS of 4.20 per USD 100 spent. This estimate positions the intervention also in the medium effective range when using LAYS as a metric scale.[30]

Overall, this section underscores that the School Leadership Professional Development Program (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.[31] 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.

7. Conclusion

This study evaluates the effects of a large School Leadership Professional Development (SLPDP) program on human capital development in Rwanda. We find no significant average effects on high stake national exams at the end of primary education. We can rule out treatment effects in the order of 0.10 standard deviations with 95% confidence. We do find an increase in test-scores in public schools of approximately 0.10 to 0.12 standard deviations, but we cannot rule out the possibility that this effect is entirely driven by chance. The rarity of published null results in the literature underscores the value of this study's findings, offering crucial insights for evidence-based policy and helping to refine expectations for program effectiveness in similar contexts.

While we do not find an effect on student test scores, we do find an improvement in the teacher retention rate by five to six percentage points. Qualitative evidence suggests further that teachers and headteachers perceive improvements in the onboarding process and teacher support following the SLPDP. However, these improvements are still deemed insufficient. Our interviews suggest that this primarily due to time constraints and resource limitations.

We suggest that future research should focus on how SLPDPs can be made more effective to improve learning outcomes. For instance, broader stakeholder involvement, additional teacher training, stronger incentives for headteachers to implement changes, or increased provision of learning resources may be required to fully unlock the program's potential. This is especially relevant given the program's promising cost-effectiveness compared to similar interventions, as discussed in Section 6.

Despite all the above, there are notable limitations and gaps in this study worth mentioning. First, this paper focus on student test-scores and teacher retention rate. Hence, measuring quantitatively other dimensions of school leadership and student well-being, teaching effectiveness and decision making in schools are some of other areas worth exploring in future researches. Second, the study examines the effect of the program in the short-term, i.e., 1–2 years after its implementation. The impact of the school leadership professional development program may manifest more prominently in the long run. Consequently, follow-up studies are needed to further comprehend its enduring effects.

[26] 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).

[27] 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.

[28] 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.

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

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

[31] 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.

CRediT authorship contribution statement

Simeon Lauterbach: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lee Crawfurd:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis. **Jocelyne C. Kirezi:** Writing – review & editing, Writing – original draft, Project administration, Funding acquisition. **Aimable Nsabimana:** Writing – review & editing, Resources. **Jef Peeraer:** Resources, Project administration, Methodology, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used CHAT GPT4 in order to rewrite paragraphs. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2025.103545.

Data availability

The authors do not have permission to share data.

References

Agirdag, Orhan, Muijs, Daniel, 2023. School leadership development and academic achievement: Effectiveness of the high performing schools programme. Int. J. Educ. Res. 122.

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

Angrist, Noam, Evans, David, Filmer, Deon, Glennerster, Rachel, 2020. How to improve education outcomes most efficiently ? A comparison of 150 interventions using thenew learning-adjusted years of schooling metric. Policy Res. Work. Pap. Ser. p. World Bank.

Asim, Salman, Gera, Ravinder Casley, Harris, Donna, Dercon, Stefan, 2024. Does Effective School Leadership Improve Student Progression and Test Scores? Evidence from a Field Experiment in Malawi. World Bank, Washington, DC.

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

Barros, de Ricardo Paes, Carvalho, de Mirela, Franco, Samuel, Garcia, Beatriz, Henriques, Ricardo, Machado, Laura, 2019. Assessment of the impact of the jovem de futuro program on learning.

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

Blimpo, Moussa P., Evans, David, Lahire, Nathalie, 2015. Parental human capital and effective school management: Evidence from the Gambia. Publisher: World Bank, Washington, DC.

Bloom, Nicholas, Lemos, Renata, Sadun, Raffaella, Van Reenen, John, 2015. Does management matter in schools? Econ. J. 125 (584), 647–674.

Branch, Gregory F., Hanushek, Eric A., Rivkin, Steven G., 2012. Estimating the effect of leaders on public sector productivity: The case of school principals. NBER Work. Pap. Ser. 17803.

Bruederle, Anna, Hodler, Roland, 2018. Nighttime lights as a proxy for human development at the local level. In: Balaguer, Jacint (Ed.), PLOS ONE 13 (9).

Burke, Mary A., Sass, Tim R., 2013. Classroom peer effects and student achievement. J. Labor Econ. 31 (1), 51–82.

Crawfurd, Lee, 2017. School management and public–private partnerships in Uganda. J. Afr. Econ. 26 (5), 539–560.

Crawfurd, Lee, 2021. Accounting for repetition and dropout in contemporaneous cross-section learning profiles: Evidence from Rwanda. Int. J. Educ. Dev. 85, 102443.

Cullen, Julie Berry, Hanushek, Eric A., Phelan, Gregory, Rivkin, Steven G., 2021. Performance information and personnel decisions in the public sector: The case of school principals. J. Hum. Resour. Publisher: University of Wisconsin Press Section: Articles.

de Hoyos, Rafael, Ganimian, Alejandro J., Holland, Peter A., 2020. Great Things Come to Those Who Wait: Experimental Evidence on Performance-Management Tools and Training in Public Schools in Argentina. Technical Report, p. 57.

de Hoyos, Rafael, Ganimian, Alejandro J., Holland, Peter A., 2021. Teaching with the test: Experimental evidence on diagnostic feedback and capacity building for public schools in Argentina. World Bank Econ. Rev. 35 (2), 499–520.

Filmer, Deon, Rogers, Halsey, Angrist, Noam, Sabarwal, Shwetlena, 2019. Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education. Policy Res. Work. Pap. 8591.

Finn, Bridgid, 2015. Measuring motivation in low-stakes assessments. ETS Res. Rep. Ser. 2015 (2), 1–17.

Gneezy, Uri, List, John A., Livingston, Jeffrey A., Qin, Xiangdong, Sadoff, Sally, Xu, Yang, 2019. Measuring success in education: The role of effort on the test itself. Am. Econ. Rev.: Insights 1 (3), 291–308.

Goldberg, et al., 2019. Effectiveness of interventions adopting a whole school approach to enhancing social and emotional development: a meta-analysis. Eur. J. Psychol. Educ. 34 (4), 755–782.

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

IIEP, 2023. School leadership. UNESCO.

JPAL, 2023. Conducting Cost-Effectiveness Analysis (CEA). Abdul Latif Jameel Poverty Action Lab (J- PAL).

King, Elisabeth (Ed.), 2013. Schooling after Genocide. Cambridge University Press, Cambridge.

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

Lassibille, Gérard, 2016. Improving the management style of school principals: results from a randomized trial. Educ. Econ. 1–21.

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

Leithwood, Kenneth, Harris, Alma, Hopkins, David, 2020. Seven strong claims about successful school leadership revisited. Sch. Leadersh. Manag. 40 (1), 5–22.

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

Lopez, Ann, Rugano, Peter, 2018. Educational leadership in post-colonial contexts: What can we learn from the experiences of three female principals in Kenyan secondary schools? Educ. Sci. 8 (3), 99.

McCord, Gordon Carlos, Rodriguez-Heredia, Mario, 2022. Nightlights and subnational economic activity: Estimating departmental GDP in Paraguay. Remote. Sens. 14 (5), 1150.

McKenzie, David, 2012. Beyond baseline and follow-up: The case for more T in experiments. J. Dev. Econ. 99 (2), 210–221.

Miller, Ashley, 2013. Principal turnover and student achievement. Econ. Educ. Rev. 36, 60–72.

MOE, 2017. Education statistics.

MOE, 2018. Education sector strategic plan. 2018–2024. Rwanda | documents | global partnership for education.

Muñoz, Pablo, Prem, Mounu, 2024. Managers' productivity and recruitment in the public sector. Am. Econ. J.: Econ. Policy.

Muralidharan, Karthik, Sheth, Ketki, 2016. Bridging education gender gaps in developing countries: The role of female teachers. J. Hum. Resour. 51 (2), 269–297.

Muralidharan, Karthik, Singh, Abhijeet, 2020. Improving public sector management at scale? Experimental evidence on school governance in India. Nber Work. Pap. 28129.

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

NISR, 2024. Home | National Institute of Statistics Rwanda.

Nzaramba, Samuel, Menon, Carl, Leonard, Phil, 2021. Gender Gaps in Secondary Schools in Rwanda: evidence from National Assessments. Report, University of Cambridge, Cambridge, p. 18.

REB, 2020. Professional standards for effective school leadership.

Romero, Mauricio, Bedoya, Juan, Yanez-Pagans, Monica, Silveyra, Marcela, De Hoyos, Rafael, 2022. Direct vs indirect management training: Experimental evidence from schools in Mexico. J. Dev. Econ. 154.

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

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

Tauchmann, Harald, 2014. Lee (2009) treatment-effect bounds for nonrandom sample selection. Stata J.: Promot. Commun. Stat. Stata 14 (4), 884–894.

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

UNESCO, 2020. Education: From disruption to recovery.

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

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

Walter, Torsten Figueiredo, 2020. Misallocation in the public sector? Cross-country evidence from two million primary schools. STICERD Work. Pap. 70.

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

Zeitlin, Andrew, 2021. Teacher turnover in Rwanda. J. Afr. Econ. 30 (1), 81–102.