NBER WORKING PAPER SERIES

MANAGING TO LEARN

Sabrin A. Beg
Anne E. Fitzpatrick
Adrienne Lucas

Working Paper 31757
http://www.nber.org/papers/w31757

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2023, revised December 2024

We gratefully acknowledge generous funding for the evaluation from the World Bank Strategic Impact Evaluation Fund, UNICEF, and J-PAL. Moussa Blimbo, Annie Duflo, Willa Friedman, Rebecca Thornton, and Sharon Wolf contributed to early stages of project development. For useful comments and suggestions, we thank Noam Angrist, Jim Berry, John Floretta, Seema Jayachandran, Chris Karbownik, Ken Leonard, Leigh Linden, Heidi McAnnally-Linz, Stephen O'Connell, Jenny Perlman and seminar participants at Emory University, the International Food Policy Research Institute, Princeton University, the University of Delaware, the University of Michigan, the University of Pittsburgh, the University of South Carolina, the University of Texas at Austin, the University of Vermont, and Vanderbilt University. For excellent research management and assistance and policy-maker engagement we thank Henry Atimone, Renaud Comba, Bridget Gyamfi, Joyce Jumpah, Edward Tsinigo, and the entire Innovations for Poverty Action Ghana team. This would not have been possible without our government partners at the Ministry of Education and Ghana Education Services. AEA RCT Registry number AEARCTR-0003977. The Innovations for Poverty Action and University of Delaware IRBs reviewed this project. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Managing to Learn
Sabrin A. Beg, Anne E. Fitzpatrick, and Adrienne Lucas
NBER Working Paper No. 31757
October 2023, revised December 2024
JEL No. H40,I25,I28,M53,M54,O15,O43

ABSTRACT

To improve public services at scale, public sector managers must encourage reticent civil servants to enact effective reforms. We show through a randomized controlled trial that when existing school managers have the tools to act as leaders, they can improve school management, student learning, and reduce teacher turnover. Additional management training differentially improved management but not student test scores. Managerial enhancements and student test score gains persisted. Our findings resolve conflicting results regarding the role of management in public sector productivity and demonstrate how public sector managers can signal reform effectiveness through personal effort, acting as leaders.

Sabrin A. Beg
University of Delaware
418 Purnell Hall
Newark, DE 19716
sabrin.beg@gmail.com

Adrienne Lucas
Lerner College of Business and Economics
University of Delaware
419 Purnell Hall
Newark, DE 19716
and NBER
alucas@udel.edu

Anne E. Fitzpatrick
Department of Agricultural, Environmental, and Development Economics
250 Fyffe Road
Columbus, OH 43210
USA
fitzpatrick.88@osu.edu

A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/3977

1 Introduction

Productivity in public service provision remains a significant challenge in developing countries, particularly in the government-run education sector. Despite millions of children attending school, many are not learning—an issue commonly referred to as the "learning crisis" (World Bank, 2017). While randomized controlled trials (RCTs) have identified interventions that can improve learning outcomes at small scales (e.g., Banerjee et al., 2007; Banerjee et al., 2017; Duflo et al., 2024), scaling these successful programs to address the crisis systemically has proven difficult. One critical barrier is the reluctance of civil servants to adopt and implement necessary reforms. Unlike in the private sector, where employee performance can be influenced by incentives like pay, promotion, or termination, public sector workers face fewer such pressures to increase their effort (Bloom et al., 2013; Bloom et al., 2020; Bruhn et al., 2018; Fenizia 2022). Yet, the public expects, and deserves, high quality public services. How can successful interventions be implemented at scale within existing public sector systems? Since public sector managers do not have the tools of compulsion, they must act as leaders, encouraging others to follow. We show through an RCT that public sector managers can act as leaders to improve student learning within existing systems and personnel, a roadmap for implementing effective programs at scale, in public sectors, where they can be truly transformative.

We partnered with the Ghanaian government to apply insights from the organizational, management, and education literatures to improve productivity at scale in the Ghanaian public education sector. In Ghana, 70 percent of elementary school students are below grade level in literacy and math (World Bank, 2018). One rigorously tested and promising solution to this learning crisis is Differentiated Instruction, a student-centered pedagogy where students are taught at their learning level instead of their grade level for part of the day (Banerjee et al., 2007; Banerjee et al., 2017; Evans and Mendez Acosta 2020; Duflo et al., 2024).[1] Yet, getting existing teachers in existing systems to implement the program

[1]Differentiated Instruction is sometimes known as Teaching at the Right Level (TaRL), Targeted Instruc-

has been elusive in Ghana and India because of the educator sector's weak managerial and oversight structures, leading to implementation rates of 6 percent in Ghana and 4 percent in India (Duflo et al., 2024; Banerjee et al. 2017). Are public sector managers the key to scaling proven interventions?

Our 210 school, three-armed RCT in Ghana demonstrates the power of managers acting as leaders to increase teacher effort and student learning. Leaders have followers, not because others must follow them but because leaders' own efforts cause followers to update their beliefs about the value of following. Public sector managers have limited management tools available to them to increase worker productivity, but they can use their own effort to update workers' beliefs about the return to undertaking more productive activities. The intervention uses this leadership framework to test two ways in which head teachers, i.e., school principals, and circuit supervisors could use their own effort to act as leaders to encourage increased teacher productivity.[2] The Teacher Training + Management Observation Treatment (Treatment 1) trained teachers in Differentiated Instruction (DI) and gave managers (i.e., head teachers and circuit supervisors) visible, costly effort tasks (primarily conducting a classroom observation and related form) that signaled the importance, value, and positive return to the new teaching methods—like leaders use their effort to elicit change from followers who they cannot compel to act. The 30 minute observation session and related form increased the time cost and thus effort required beyond status quo monitoring where head teachers would peek into a classroom, attempting to remain unseen, or count completed pages in students' exercise books. The intervention was not monitoring to ensure control and compliance, as no rewards or punishments occurred, but instead about influencing behavior through a demonstrated commitment of effort. The Teacher Training + Management Observation + People Skills Treatment (Treatment 2) included all the elements of Treatment 1 plus an additional training for school managers on "People Skills" to increase their mentoring

tion (TI), or Differentiated Learning (DL).

[2]Circuit supervisors oversee 8 geographically proximate schools and report to the District Education Office. A similar managerial layer exists in other contexts with historically challenging communication over distances.

and support, improving sometimes fraught interpersonal relationships between teachers and their managers.[3] Using the improved People Skills was an another form of manager effort but directed at the interpersonal relationships not signaling the value of undertaking specific classroom pedagogy. The final group was the control group that continued with business as usual. All material design and distribution and training was implemented by existing civil servants under the Ghanaian Ministry of Education umbrella.

We find that the two interventions resulted in increased head teacher effort and improved school management, classroom teaching quality, and student learning. The gains persisted after the end of the intervention. Both treatments increased teacher engagement with their classrooms and led to equal, high Differentiated Instruction implementation, ten and fifteen times higher than versions without increased manager effort (Banerjee et al., 2016; Duflo et al., 2024). Both treatments increased the quality of Observational Practices by 0.4 standard deviations (SD) relative to the control group. The additional mentoring and support training in the People Skills arm improved the quality of People Skills relative to the other two arms.[4] The interventions further improved school norms around shirking, something hard to alter and essential for sustained change (Coch and French 1948; Tankard and Paluck 2016). In the year after the intervention, treatment schools had higher management quality, less teacher turn-over (50 percent less in the Teaching Training + Management Observation intervention), and were still implementing DI without continued training or refreshed materials.

The differential improvements in People Skills but similar changes in Observational Practices and classroom activities led to the same 0.11 SD improvements in test scores (about one third of a year of schooling), similar in magnitude to the DI implementations that included NGO personnel as leaders (Banerjee et al., 2017). Students from treatment schools continued to have test scores that were higher than their control school peers two years after

[3]In the private sector, People Skills can also include elements of personnel management like wages and hiring and firing. None of those aspects are controlled at the local level in Ghanaian schools and are not part of our intervention.

[4]These measures of management quality are adapted from the internationally validated Development World Management Survey (D-WMS) competencies (Bloom et al., 2015; Lemos and Scur, 2016). Management of schools includes more than Observational Practices and People Skills. See more details on management measurement in Section 5.3.

the end of the intervention, through COVID-19-related school closures and students graduating from primary to junior high school. Thus, management effort induced higher teacher implementation than those achieved through teacher training alone in prior interventions, pointing to the important role of managers and Observational Practices as complementary to teacher effort.

The importance of Observational Practices and not People Skills to increase student test scores is further supported by a mediation analysis: increases in student test scores were positively related to increases in Observational Practices and unrelated to increases in People Skills. Student learning only improved when what was happening in the classroom changed–managers acting as leaders induced classroom change but additional changes to interpersonal relationships did not further improve test scores because they did not alter teaching effort. This provides insight into contradictory findings on the importance of management in other studies.

Our results are consistent with the conceptual framework of leadership—teachers increase their effort when managers' actions credibly signal the return to that effort, whether productivity (Hermalin 1998) or as an organizational value (Spence 1973; Gibbons and Roberts 2013). Improvements to productivity, in our context student learning, will only occur if teachers' increased efforts directly affect students. Any other managerial activities, such as additional support and mentorship, will only affect productivity if they lead to additional worker effort.

In addition to its policy relevance, as national governments in Africa and state governments in South Asia are implementing similar programs that could reach millions of children at scale, our paper makes three related contributions to the economics literature. First, we show the potential to transform the provision of public services by giving public sector managers tasks that encourage workers to increase their level of effort–causing managers to act like leaders in instigating change. The existing literature on leaders focuses on how someone who is already a recognized leader influences others (theory: Hermalin 1998; Loeper et al.,

2014; Acemoglu and Jackson 2105; Akerlof and Holden 2016; empirical: Boudreau et al., 2022; List and Lucking-Reiley 2002; List and Rondeau, 2003; Andreoni, 2006; Potters et al., 2007; Brandts et al., 2016; Koukoumelis et al., 2012). Principals and circuit supervisors, or public sector managers generally, are inherently weak in the management tools at their disposal, are not necessarily leaders, and not selected based on their leadership ability (Fenizia 2022; Lucas 2024). They can proclaim priorities but workers are not obligated to follow, instead they have to convince workers to update their belief that exerting effort is in their interest.[5] School mangers increasing their Observational Practices demonstrated to teachers the value of Differentiated Instruction, with persistent effects one year after the program ended. The importance of mangers as leaders is a blueprint for scaling other successful NGO or researcher-initiated programs in the public sector within existing systems where truly transformative change will take place.

Second, we remedy a recent puzzle in the literature on management's relationship with productivity—why do improvements in management not necessarily increase productivity? We increased both management quality and test scores, while other recent papers increased management quality but not productivity (Romero et al., 2022; Muralidharan and Singh 2020; Hoffman and Tadelis 2021; Ganimian and Freel 2020; de Hoyas et al. 2019).[6] The breakdown in relationship between management and student learning is in contrast to previous correlations from the private and public sectors, and evidence on management improvements increasing productivity in the private sector.[7] Improving management in the public sector in developing countries is a nascent field with no papers appearing prior to 2017 (Finan et al., 2017).[8] Outsourcing school management to private sector firms (Romero, Sandefur,

[5]Our use of an observation form that leaders can use to signal their own belief in a program broadens the literature on checklists into schools (e.g., Haynes et al., 2009; Semel et al., 2010; de Villiers 2013; Oliver et al. 2015; Choi et al., 2016; Gray-Lobe et al., 2022).

[6]de Hoyas et al. (2020) found positive effects on student passing rates and reduction in failure rates two years after the conclusion of a two year performance management intervention.

[7]For private sector correlations see, e.g., Bloom and Van Reenen (2007); Hsieh and Klenow (2009); summary in Bloom and Van Reenen (2011); Bloom et al., (2019); and Gibbons and Henderson 2012. For public sector correlations see Rasul and Rogger (2018); Tsai et al. (2015); Bloom et al., (2015); Lemos and Scur (2016); Crawfurd (2017); Lemos et al. (2021). For evidence on management improvements in the private sector see Bloom et al., (2013); Bloom et al., (2018); and Bruhn et al., (2018).

[8]The evidence in developed country schools is scant and often involves substantial (300 hour) trainings

and Sandholtz 2020) or creating additional management layers parallel to the existing system (Cilliers et al. 2020a; Cilliers et al. 2020b) improved student learning, leaving open questions of the most important pieces of management to improve within existing systems. By separating management into multiple components, we show that not all management is equally important in the production of learning. Managerial practices increase learning when they improve worker effort—better Observational Practices changed classroom activities and increased learning while additional improvements in People Skills did not further augment either. We also show that existing public sector school management can be improved and can increase student learning.

Third, we show that governments have the potential to scale effective programs. This program was designed and implemented solely by the government, and resulted in substantive increases in student learning, persistent gains in management, and reduced the vexing problem of teacher turnover. Governments expanding access to effective programs is neither trivial nor obvious because of bureaucratic inefficiencies and limitations to state capacity (Muralidharan, Niehaus, and Sukhtankar 2016; Muralidharan and Niehaus 2017; Bold et al. 2018). Across studies, government-implemented programs have smaller effect sizes than those implemented by academics or NGOs (Vivalt 2020). The embrace and integration of this program by the education system, instead of it being an outside imposition, created a strong foundation for success.

Overall, the interventions increased management and student learning during the year of the intervention, management for at least one year later, and student learning for at least two years, even after graduation to junior high school. Starting in 2022, the Teacher Training + Management Observation intervention was being scaled to 10,000 schools in Ghana.

(Fryer 2017).

2 Background

As with many other countries, Ghana is beset with the dual challenge of heterogeneous classrooms and low average student achievement (Glewwe and Muralidharan 2016; Ministry of Education 2018; Duflo et al., 2024). Primary schools in Ghana are grades 1 (P1) through 6 (P6). In primary schools, teachers are classroom teachers, teaching all subjects to their assigned students. After primary school, students continue to junior high school (JHS) grades JHS1 through JHS3, similar to grades 7 through 9 in the US. JHS often combine students from multiple primary schools. Our study sample are students in grades 5 and 6 at the start of the intervention. These students should have been in JHS1 and JHS2 at our final data collection point. Our study straddles the COVID19-induced 10 month Ghanaian school closures and school calendar shift from a September start in 2019 to a January start in 2021.[9]

Primary school teachers are employed, assigned to schools, and transferred between schools by the national Ghana Education Services (GES). A head teacher, i.e., school principal, is the manager of each school but does not control the terms of teacher employment.[10] Each school belongs to a circuit of approximately 8 geographically proximate schools, overseen by a circuit supervisor.[11] Circuit supervisors act as liaisons between the school and the District Education Office but have no discretion over the terms of employment of teachers or head teachers. As part of typical operations, both circuit supervisors and head teachers observe teaching, usually a 5 minute peek through a door or window.

[9]School years in Ghana consist of three terms. The intervention year (study students in grades P5 and P6) was a normal September 2018 to July 2019 school year. The year after (study students in grades P6 and JHS1), the school year started September 2019 and abruptly ended in March 2020 due to COVID-19. The next school year (study students in grades JHS1 and JHS2) started in January 2021. Ghana Education Services provided limited remote learning opportunities through television broadcasts during the closures. See Fitzpatrick et al. (2021) for respondents' experiences with COVID-19 beyond their school duties.

[10]Head teachers are almost always previous classroom teachers. Most do not have any additional formal training prior to becoming a head teacher, and 28 percent of our sample did have any head teacher-specific training prior to this intervention.

[11]As with head teachers, circuit supervisors are almost always previous teachers and, like head teachers, receive almost no additional formal training. Since the conclusion of this study, the Circuit Supervisor title has been replaced by the title School Improvement and Support Officer (SISO).

3 Intervention

3.1 Differentiated Instruction

Differentiated Instruction (DI), also known as Differentiated Learning (DL), Targeted Instruction (TI), or Teaching at the Right Level (TaRL), is an active, student centered pedagogy that teaches students at their learning level instead of their grade level for part of each school day or in lessons outside of school time. As of 2024, it was part of education policy in sixteen countries across Africa and South Asia. It improves test scores when educators implement it (Angrist and Meager 2023). The most successful implementations of DI involved the use of assistants or support from non-governmental organizations. In contrast, a teacher implemented version in Ghana had implementation rates of only 6 percent and no increase in student learning after one year (Duflo et al., 2024). Similarly, existing teachers with existing supervisory structures only implemented grouping by ability during 4 percent of observations in India (Banerjee et al. 2016). Could successes be replicated with existing personnel?

Both interventions were developed in partnership with government stakeholders to serve as scalable solutions for implementing DI within existing systems—a goal that had previously proven elusive (Banerjee et al., 2007; Banerjee et al., 2017; and Duflo et al., 2024). For this version of DI, existing teachers were to divide students across three grade levels by three learning levels (instead of their grade levels) for part of the school day. The success of previous implementations with personnel outside the existing system pointed to management as a potential key for implementation at scale. Yet, any management intervention had to work within the constraints of the public sector system where the typical levers of coercion in the private sector, such as the terms of employment, are unavailable. This intervention focuses on two aspects of management that overlap with leadership—exerting effort in observing production and interpersonal relationships—to encourage teachers to implement Differentiated Instruction. This sub-sections describe the two interventions. The initial trainings

occurred prior to the start of the 2018-2019 academic year. All trainings and materials were government designed and delivered.

3.2 Treatment 1 – Teacher Training + Management Observation

Treatment 1 trained teachers and provided tools and opportunities for managers to use their own effort to change teachers' beliefs about the positive return to implementing the differentiated instruction program in schools.

Teacher Training

Teachers of grades 4 through 6 received in-service training on how to implement Differentiated Instruction, i.e., dividing their students by learning level instead of grade level, and teaching and learning materials to support a more student focused, active learning practice, the same partial day tracking as in Duflo et al., (2024).[12] Teachers had discretion about how much effort to exert on this new, to them, teaching method. Superior pedagogy will only increase student learning if it is implemented in the classroom. Teachers would be more likely to implement a program that they believe to have a positive return, whether to their students' learning, how they or their school is perceived by other actors in the system, or their future career prospects.

Management Observation

One of the few tools available to public sector managers to increase the productivity of their workers is the spending of their own effort. In the Management Observation arm managers used their own effort to observe production and highlight the important aspects of DI that increase student learning.

In Ghana, and many other countries, teachers are bombarded with trainings and pronouncements of what they should and should not be doing in their classrooms. Taking on

[12]Students in treatment schools were to have DI time with their learning level one hour each day, four days per week, eight weeks per term. One teacher was assigned each learning level. Because learners from all three grades were in each level, the three teachers had to work together–if one teacher was absent or did not want to participate, the students of that level would have no where to go during DI time and DI could not happen. See Section A.1 for more details on Differentiated Instruction and the exact implementation.

any new method requires effort. When the returns are uncertain or perceived to be low, teachers will not exert that effort. Unlike in the private sector, head teachers and circuit supervisors cannot adjust employment conditions to reflect teacher effort. Instead, they can change teachers' perceptions of the return to effort through their own actions.

This intervention embedded multiple opportunities for head teachers and circuit supervisors to "lead by example" by spending their own effort to update teachers' beliefs about, or signal, the value of implementing the revised teaching methods. The most costly effort from managers, and likely the most salient signal to teachers, was the head teachers' and circuit supervisors' 30 minute classroom observation and corresponding use of a classroom observation form that emphasized the core components of DI–dividing students by learning level, teaching at the level of the student, and engaging in active, student centered pedagogy. Head teachers and circuit supervisors were instructed to use this observation form at least once per term with each teacher, providing them a straightforward task that cost them effort and signaled to teachers that this program was a priority and valued by the system. Head teachers were already observing teachers, but doing so by peeking in a window for 5 minutes and using an evaluation form that focused on whether students completed exercise books and were well behaved. This intervention was designed to increase the time managers spent observing worker productivity and focused observations on documenting and measuring aspects of teaching that increase student learning.

As an additional display of effort, head teachers and circuit supervisors attended the teacher training, which also helped them learn how teachers should implement DI and support teachers in that implementation. Interacting with district and regional trainers demonstrated to the head teachers and circuit supervisors that this program was a priority. Finally, a national level monitoring team displayed their own costly effort by visiting all study districts each term to observe the intervention, providing an additional credible signal to the teachers, head teachers, circuit supervisors, and district education officials that this intervention had the support of the national leaders in education.[13]

[13]National level monitors visit schools as regular policy, but less frequently, averaging about 16 percent of

3.3 Treatment 2 – Teacher Training + Management Observation + People Skills

Treatment 2 received the same Teacher Training + Management Observation intervention as in Treatment 1 plus additional People Skills training.

People Skills

Improving interpersonal relationships is another way in which managers can spend their effort. This intervention provided managers tools to improve these relationships. Head teachers and circuit supervisors received additional training, a handbook, and quick reference cards on how to be effective supporters of teachers, e.g., how to make a constructive criticism sandwich by framing an area of weakness around two compliments, and a "help desk" number that they could call or text with questions for support.[14] Unlike the task-based observation form in Treatment 1, these trainings and materials were more holistic and sought to transform the relationship between teachers and their two immediate layers of supervisors into one of collaborative support for improved learning. During the implementation year, head teachers received automated, supportive, weekly text messages to remind them of key aspects of a supportive relationship with teachers and differentiated instruction components and dates.[15] Whether improved manager-worker relations induce teachers to increase their own effort, are valued by teachers but do not change effort, or are something about which teachers are indifferent are empirical questions.

schools per year. During the intervention, they visited 88 percent of treatment schools.

[14]Despite head teachers, circuit supervisors, and other stakeholders insisting that a help-desk was important, the help-desk received almost no calls or texts during the intervention. Even though it was barely used, its existence could have further signaled that national stakeholders were exerting effort towards implementation.

[15]For example, "Remember to compliment each teacher this week" or "This is the week for student leveling."

4 Empirical Strategy

The primary conceptual difficulty in estimating the effect of training and policies on management quality, classroom activities, and student outcomes are the endogenous nature of training and policies and their correlation with other unobserved aspects of the school. To overcome this difficulty, we conducted a randomized controlled trial. We randomized 210 schools into one of three treatment arms: Teacher Training + Management Observation (T1), Teacher Training + Management Observation + People Skills (T2), and control.[16] From this randomization design, estimation of treatment is straightforward. Specifically we estimate,

$$y_{is} = \beta T1_s + \beta_2 T2_s + \Gamma' X_{is} + \varepsilon_{is} \tag{1}$$

where y_{is} is outcome y for respondent i in school s, $T1_s$ is an indicator variable equal to one if the school was in the Teacher Training + Management Observation treatment, $T2_s$ is an indicator equal to 1 for schools in the Teacher Training + Management Observation + People Skills treatment, X_{is} are a vector of school and individual level controls including strata (district) fixed effects and baseline level y_{is} as appropriate, and ε_{is} is a cluster-robust error term assumed to be uncorrelated between schools but allowed to be correlated within a school.[17]

We test the impact of the treatments on school management practices and norms, classroom activities, program implementation, and students' test scores, attendance, and persistence in school. We compare the effect of Teacher Training + Management Observation relative to the status quo through the estimate of β_1. The estimate of β_2 captures the total

[16] We do not include an intervention that is teacher training alone as both Duflo et al., (2024) in Ghana and Banerjee et al., (2017) in India found statistically insignificant increases in test scores one year after similar training when the program was not accompanied by any changes to management or supervision.

[17] In student-level regressions, to improve precision we additionally control for student age and age-squared, baseline grade level, an indicator variable for being female, and baseline test score for test score regressions. In teacher-level regressions, we additionally include teacher age, age-squared, years of experience, years of experience-squared, gender, and baseline class. In specifications with multiple waves, we include an indicator for the survey round. Regressions with outcomes on classroom level implementation include the average percent of teachers present during the baseline survey.

effect of Teacher Training + Management Observation + People Skills relative to the status quo. The difference between the estimates of $\beta_1$and β_2 provides the differential effect of People Skills. We provide further insight into the relative and absolute importance of the two interventions through a mediation analysis.

5 Sample Selection and Data

In this section we first describe how we constructed the sample and then the data collected.

5.1 Sample Selection and Randomization Procedure

This study occurred in 210 schools across 20 districts spread across eight of ten regions of Ghana.[18]

Figure 1 contains the study design. From the 140 study circuits, effectively the universe of circuits in the study districts, we randomly assigned 70 circuit supervisors to receive the Management Observation intervention and 70 circuit supervisors to receive the Management Observation + People Skills intervention.[19] [20] We randomly selected two eligible schools from each circuit.[21] Within each of the Management Observation circuits, we randomly

[18]The districts were between the 77th and 1st (poorest) percentile of the national district level poverty ranking. These districts were selected because the United Nations Children's Fund (UNICEF), one of the implementation funders, had existing relationships in these districts.

[19]The 20 districts contained 145 circuits. Ex ante we excluded two circuits from the study for piloting and three circuits because they had too few eligible schools, leaving 140 study circuits. See below for details on school eligibility.

[20]Circuit supervisors of control schools received the Management Observation intervention. Therefore, our estimates are lower bounds of the overall effect size relative to a pure control school. The effect of a trained circuit supervisor on an otherwise untreated school is likely small. They were instructed not to use the DI observation form in non-DI schools, and we did not observe any control schools dividing their students by learning levels. As the previous versions of DI were barely implemented and the methods did not transfer even between educators in the same school, spillovers are unlikely (Duflo et al., 2024). If spillovers happened, they will bias our results towards 0. The study design was based upon ensuring adequate statistical power for the student test score outcome given the total number of circuits in the 20 districts, the available budget, and estimates from previous studies.

[21]Schools from the union of the official Education Management Information System (EMIS) and UNICEF rosters of schools that operated only a single shift per day and had positive enrollment and separate sections for P4 through P6 were eligible. Schools with shifts have altered time-tables that would not be amenable to differentiated instruction time, and schools with multiple grades taught in a single classroom would not have enough teachers to separately teach three learning levels. We contacted schools to confirm these

assigned one selected school to be a control school and the other to be a Teacher Training + Management Observation school. To be consistent across treatments, within each of the Management Observation + People Skills circuits, we randomly assigned one selected school to receive the Teacher Training + Management Observation + People Skills intervention and removed the second school from the study. Our sampling strategy ensures that the circuit and school selection is identical in the three groups.

[Figure 1 about here]

The resulting experimental sample is 210 schools across three study arms: Teacher Training + Management Observation, Teacher Training + Management Observation + People Skills, and control.[22] This scheme results in two treatment arms and one control arm, with 70 schools in each arm. Our design has the advantage of allowing us to measure the impact of Teacher Training + Management Observation compared to business-as-usual, and the impact of Teacher Training + Management Observation + People Skills compared to business-as-usual. It also allows us to compare the differential effect of adding the People Skills intervention relative to the Teacher Training + Management Observation intervention alone. In all cases we establish both effectiveness and cost-effectiveness.

5.2 Data Collection

To evaluate the effect of the two interventions, we conducted five rounds of data collection across three years – a baseline prior to implementation, two spot checks during the implementation year, a follow-up in the final term of the implementation year, a truncated (due to COVID-19 school closures) spot check round the year after the intervention, and a follow-up

characteristics.

[22]This study is similar to a fully cross-randomized design, although we do not have any schools that only received the People Skills training without Teacher Training + Management Observation nor schools that received only Teacher Training. We follow Muralidharan et al. (2019) and use previous research to exclude and rule out these ineffective intervention arms. Previous research has shown the ineffectiveness over a one year intervention horizon of Teacher Training alone (Banerjee et al., 2017 and Duflo et al., 2024) and Management training alone (Muralidharan and Singh 2020; Ganimian and Freel 2021; Romero et al., 2022).

two years after the end of the intervention. The full project timeline appears in Figure 2 with school calendar dates above the line and study activities below the line.

[Figure 2 about here]

Baseline

To ensure a baseline prior to anyone receiving treatment, the baseline occurred in May and June 2018, near the end of the 2017-2018 academic year.[23] We surveyed the circuit supervisors and the head teachers and P4 through P6 teachers in the 210 study schools on their backgrounds and existing management and teaching practices. Additional details on the management measures appear in Section 5.3. We further surveyed and tested a random sample of 15 students from P4 and 15 students from P5.[24] The student assessments included both foundational and grade level content. Additional details on test construction appear in Appendix Section A.2. We compute an overall test score using item response theory and standardize based on the baseline mean and standard deviation. We surveyed and tested 5,893 students at baseline.

Spot-checks

We conducted three spot-check visits, two during the 2018-2019 school year (one in Term 1 and one in Term 2) and one during Term 2 of the 2019-2020 school year. During each spot-check visit, enumerators arrived unannounced and recorded the attendance of the head teacher, teacher, and baseline students. Circuit supervisors, head teachers, and teachers responded to surveys on management and program take-up and implementation. We also conducted two classroom observations during the first two periods of the day in each P4 through P6 classroom, noting teacher presence, whether students were divided by grade level instead of learning level, and the use of teaching and learning materials. We further collected basic demographics on any new head teachers or teachers since the baseline. All

[23]The teacher, head teacher, and circuit supervisor training occurred in August 2018, before the start of the 2018-2019 academic year.

[24]If a school had more than one section of a grade, we first randomly selected a section then selected students from that section. If a school had fewer than 15 present students in a grade, we surveyed and tested all students who were present in that grade.

schools should have been visited in each spot check, but the third spot check was disrupted by the Covid pandemic.[25]

Achievement Follow-ups

We conducted two follow-up survey rounds, one at the end of the 2018-2019 school year (students should have been in grades P5 and P6) and one in the middle of the 2020-2021 school year (students should have been in grades JHS1 and JHS2). During these visits, we attempted to survey and invigilate exams for all baseline students, following up at other schools and homes as necessary. These tests were similar to those at baseline, but included additional, harder questions with some of the easier questions removed. In addition, we conducted interviews of circuit supervisors, head teachers, and teachers from the original primary schools. These surveys collected information on school enrollment, organization and management, teacher support, mentorship, and program implementation.

5.3 Management Indices

We measure the effect of the interventions on three measures of management. The first two, Observational Practices and People Management, reflect the core management foci of the two interventions. The third, Other Management, measures other aspects of management that were not emphasized in the intervention but could be affected as managers become more engaged with their schools. We combine teachers', head teachers', and circuit supervisors' responses and enumerators' direct observations into a single metric for each index following Anderson (2008). Each index is standardized relative to a control group mean of 0 and standard deviation of 1. Our measures include components of the previously validated Development World Management Survey (D-WMS) (Lemos and Scur 2016). The specific questions and precise mapping appears in Appendix Table A1.

Observational Practices

[25] We stopped the in-person portion of the third spot-check in March 2020 because of Covid19 transmission, having reached only 60 percent of our sample schools. We surveyed the remaining teachers and head teachers via phone but could not perform classroom observations.

One of the core functions of management is observing employee performance relative to organizational standards. The interventions encouraged head teachers and circuit supervisors to observe classroom teaching and record their thoughts on an observation form that highlighted the key components of Differentiated Instruction, assisting the head teachers and circuit supervisors in performing this core function and sending the costly effort signal to teachers of the value and importance of Differentiated Instruction. The Observational Practices index includes reports from teachers, head teachers, and circuit supervisors about these observations.

People Skills

One of the few other management levers available to public sector managers are their interpersonal relationships with employees. The second intervention included People Skills training and tools designed to improve these relationships. The People Skills Index includes responses from teachers, head teachers, and circuit supervisors related to how supervisors and subordinates interact, the dialogue around performance review, the relationship between subordinates and supervisors, and the degree to which teachers received mentoring, coaching, and feedback to improve their teaching practices.

Other Management

The final index, Other Management, includes other aspects of management, which were not targeted by either intervention, but might be hallmarks of highly functioning school management, such as monitoring student attendance, continuous improvement, and consequence management. The interventions did not specifically target these aspects of school management, but they could have changed as head teachers became more engaged with their schools.

5.4 Summary Statistics and Baseline Balance

Appendix Table A2 displays the summary statistics and tests for baseline balance across the three treatment arms for Student (Panel A), Teacher (Panel B), Head Teacher and School

(Panel C), and Circuit Supervisor (Panel D) variables. Across all means tested, none of the differences are statistically significantly different from 0.

A few statistics, which are statistically equivalent across all arms, provide useful context. Students were on average about 12 years old, about 2 years older than than expected if students started school on time and continued apace. About 53 percent of students, 74 percent of teachers, 80 percent of head teachers, and 90 percent of circuit supervisors were male. Teachers were about 31 years old, head teachers were about 42 years old, and circuit supervisors were 45 years old. Almost all teachers (91 percent) had a bachelor's degree or a diploma, and 28 percent of head teachers had received no pre-service or in-service training specifically about being a head teacher.

6 Results

We first estimate the effects of the interventions on management quality. We then test if school and classroom operations changed, as management practices only improve learning if teacher effort on classroom practices change. Next, we test for the downstream effects of the interventions on students' cognitive and non-cognitive outcomes. Finally, we estimate the persistence of the effects in the two years after the conclusion of the intervention.

6.1 Effort Signal and Management Practices

Both interventions increased management quality (Table 1). Observational Practices were about 0.4SD better in treatment schools, relative to the control group index standardized to mean 0 and standard deviation of 1, statistically indistinguishable across the two treatments (column 1). The effects of the interventions on the individual components of the Observational Practices Index appear in Appendix Table A3.

[Table 1 about here]

The effect of the interventions on People Skills is only statistically significant for the inter-

vention that included specific, additional training on People Skills, increasing that measure by 0.5SD (column 2). We reject the equality of the coefficients across the two interventions. The effects of the interventions on the individual components of the People Skills Index appear in Appendix Table A4.

The final column of Table 1 tests for the effects of the interventions on Other Management – aspects of management not directly related to the intervention but which could have improved as head teachers became more interested and engaged in their schools and the school climate changed (see more details on the changing of school norms below). These aspects increased in both treatment arms—about 0.4SD for the Teacher Training + Management Observation and 0.6SD for the Teacher Training + Management Observation + People Skills—with the effect of the intervention that included People Skills statistically larger at the 10 percent level (p-value=0.07; column 3). Appendix Table A5 contains the estimates of the effect on each component of the Other Management index separately.

Therefore, managers in both treatments used costly, credible effort to increase their Observational management practices. These improvements were statistically equivalent across the two interventions. The schools that received the additional People Skills training had higher People Skills quality than both the control and the Teacher Training + Management Observation schools. Both interventions further had spillovers into other aspects of management, more so for the intervention that included People Skills.[26]

6.2 School Operations, Classroom Activities, Norms, and Stress

Operations

Improvements in management practices can only affect student outcomes if school is in session and head teachers and teachers are present. To test for changes in school operations,

[26]The results in Table 1 are not respondents overstating their own actions—the effects are similar, including the differential effects for the People Skills arm, when the indices remove the self-reported outcomes and rely only on subordinate reports (Appendix Table A6). Teachers (the subordinates) received the same training in both interventions. The interventions increased a composite measure of management by 0.3SD (no People Skills training) to 0.4SD (with People Skills training, Appendix Table A6).

we conducted two unannounced spot checks when schools were supposed to be holding normal school operations and DI lessons. The two interventions were equally effective at improving school operations—the interventions increased the likelihood that school was in session by 3 percentage points and head teachers were present by about 11 percentage points (Appendix Table A7).[27]

Classroom Activities

The interventions improved classroom instruction as teachers implemented the practices from the teacher training (Table 2). Teachers were more likely to be in the classroom (11 percentage points, column 1) and their teaching involved more active learning (column 2), techniques taught in the Teacher Training.[28] The most active teaching was in Teacher Training + Management Observation schools with less in Teacher Training + Management Observation + People Skills schools, and even less in control schools. Focusing on their additional People Skills training might have distracted head teachers from these broader classroom practices.

[Table 2 about here]

In addition to using active pedagogy, the core of DI is to divide students by learning levels for one period each day. We observed about 60 percent of schools teaching their students by learning level instead of grade level during at least one of first two periods of the day of our unannounced visits (column 3).[29] Relative to implementation without manager involvement, this level of implementation is ten times previous implementation rates in Ghana and fifteen times previous implementation rates in India (Duflo et al., 2024; Banerjee et al. 2016). Therefore, the engagement of managers appears key to implementation.

Column 4 combines this enumerator observed outcome with additional self-reported measures of DI implementation into a single index finding that the average school across both

[27]We do not find any effect on teacher attendance (Appendix Table A7). About half of this absenteeism appears to be chronic–being absent at the first spot check increased the likelihood of being absent at the second spot check by 15 percentage points (p = 0.01), making it difficult to change.

[28]The effects on the individual Active Learning Index components appear in Appendix Table A8 Panel A. The results are similar if we condition on a teacher being present in the classroom (Table A8, Panel B).

[29]As some schools reported doing DI later in the day, 60 percent is likely an understatement of adherence.

treatment arms was implementing 80 percent of the program.[30] Teachers exerted effort in implementing DI.[31]

Norms and Stress

The interventions decreased the acceptability of shirking. We provided teachers and head teachers vignettes in which a hypothetical student, teacher, head teacher, or circuit supervisor was shirking, e.g. managers not providing useful feedback or someone being chronically absent, and asked the respondents whether the behavior was acceptable. In Table 3 we estimate Equation 1, replacing the dependent variable with an index, standardized by control mean and standard deviation, over the vignettes about a specific agent. A higher value indicates that the shirking behavior was more acceptable. Teachers in treatment schools judged shirking behavior by hypothetical teachers more negatively by about 0.3SD (column 1). They judged shirking by all school agents, an index that combines responses about students, teachers, head teachers, and circuit supervisors, more negatively by 0.3SD (Teacher Training + Management Observation) or 0.4SD (Teacher Training + Management Observation + People Skills) (column 2). Head teachers similarly viewed shirking by teachers about 0.3SD more harshly (column 3). The intervention did not change the overall opinions of head teachers, just their views of shirking by teachers (column 4).[32]

[Table 3 about here]

The additional effort by head teachers in Table 1 and teachers in Table 2 and higher standards in Table 3 did not lead to increased stress and burnout for teachers in the Teacher Training + Management Observation intervention or head teachers in either intervention

[30]The additional index components are whether the school completed the leveling, did DI at least four days in the prior week, had done or were planning to do DI the day of the enumeration visit, and still had their copy of the DI manual at our final visit. Appendix Table A9 shows the effect of the interventions each component of the index—in all cases the point values are larger for the Teacher Training + Management Observation + People Skills intervention but only statistically different (p-value=0.06) for completing the leveling.

[31]In Appendix Table A10 we estimate the outcomes in Table 2 separately by Period as we observed both Period 1 and 2. The improvements relative to the control group are present in both Periods. The differential effects between the two treatments for the Active Learning Index are only in Period 2, when schools were less likely to be engaged in DI.

[32]Appendix Table A11 presents the teacher and head teacher opinions about the other agents at the school separately.

(Appendix Table A12). Teachers in the Teacher Training + Management Observation + People Skills intervention reported higher levels of stress and burnout than the control group perhaps because head teachers' implementations of improved interpersonal relationship training felt like additional monitoring instead of support.

6.3 Student Outcomes

Achievement

We estimate the effect of the two interventions on student achievement using Equation 1 with a student's endline score as the dependent variable, including their baseline test score as an additional covariate. Table 4 contains these results. Students in either treatment increased their overall test scores relative to the control group by 0.11SD (column 1) with 0.13SD improvements in Math (column 2) and 0.07SD improvements in English (column 3).[33] Over this same period, one school year, control group students learned about 0.3SD—the interventions provided an additional one third of a year of learning with one academic year of exposure. The point values across the two interventions are statistically equivalent with point values within 0.01SD of each other. Therefore, even though the quality of People Skills differentially changed in the intervention that included People Skills training, the test score improvements were almost identical across the two interventions.[34]

[Table 4 about here]

We tested for heterogeneity in program impacts on students' test scores by students'

[33] Appendix Table A13 provides additional estimates of the effects on student test scores. Panel A limits the controls to only the baseline test score and strata, finding similar estimates. Panels B-D estimate the effect on subsets of questions of different difficulty levels, with larger effects for foundational questions. The increase in test scores is likely not only the result of increased teacher time in the classroom based on existing estimates of the relationship between teacher attendance and student test scores (Das et al., 2007; Duflo et al., 2012; Herrmann and Rockoff 2012; Gershenson 2016; Cilliers et al., 2018).

[34] As with any RCT, one concern is non-random attrition at the follow-up generating differential selection into the test by treatment status. To limit attrition we tracked all students not present in school at the start of the follow-up visit, eventually testing 96 percent of all baseline students. We tested 1.5 percentage points fewer students from the Teacher Training + Management Observation + People Skills arm than the control group but this is not differential by both test score and treatment (Appendix Table A14). Nevertheless, in Appendix Table A15 we provide Lee (2009) bounds accounting for this marginally differential attrition. The point values are similar with the same statistical significance as those with the full sample.

baseline test score, gender, grade level at baseline, parents' literacy, and a school amenity index, finding no heterogeneous effects (Appendix Table A16).[35] Appendix Figure A1 displays the non-parametric effects on test scores by baseline score. Effects are approximately equal throughout the distribution of baseline test scores, similar to the findings in other studies in which material was provided for multiple learning levels, not only focused on remedial learners (e.g., Banerjee et al. 2007; Banerjee et al. 2017; Duflo et al., 2024).

Non-cognitive Outcomes

Dividing students by learning levels might affect students' motivation, which could manifest itself in students' non-cognitive outcomes such as absenteeism, drop out, opinions about school, and future schooling aspirations. We find at most minimal effects on absenteeism and drop-out (Appendix Table A17, columns 1-3).[36] The interventions did not decrease enthusiasm about school or future aspirations (columns 4-6). Instead, the Teacher Training + Management Observation intervention increased the likelihood that students reported liking English or Math, i.e., the two intervention subjects, very much by 5 percentage points, statistically different from both the other intervention and the control group. Overall, we find no evidence of student discouragement.

[35]We tested for heterogeneity by test score as students who were the most behind grade level might have benefited more, by gender because of evidence from Ghana that head teachers are systematically biased against female teachers (Beg, Fitzpatrick, and Lucas 2021) and therefore might also exhibit bias against female students, by student grade level as students at different places in their scholastic journeys might have had different experiences, by parent literacy as a proxy for whether the student was a first generation learner and could get help at home if needed, and finally by school amenities in case the intervention was particularly well suited for schools with different levels of existing resources.

[36]Students in the Teacher Training + Management Observation intervention were about 3 percentage points more likely to be absent during the spot-check, but their schools were about 3 percentage points more likely to be open (and we could not check attendance in closed schools), therefore the net effect relative to the control group is approximately 0 (Appendix Table A17, column 1). The interventions increased by about 2 percentage points the likelihood that the teacher reported the student was no longer enrolled in the school, the sum of drop-outs and transfers (column 2). The interventions may have increased teachers' awareness of students who were no longer attending their school since the intervention instructed teachers to to test and record scores for students each term. When reached directly, students across all treatment arms were equally likely to report that they were still attending school (column 3).

6.4 Persistent Effects

We test the effects of the interventions one year after the end of the intervention on schools and two years after the end of the intervention on students—finding persistent effects on both.

Management

Management improvements persisted into the year after the intervention.[37] Observational Practices were still superior to those in the control group in the Teacher Training + Management Observation + People Skills schools (Table 5, column 1). People Skills were statistically significantly higher in both types of treatment schools than control schools (column 2).[38] Across both management indices, the intervention that included People Skills training had larger, but not statistically different, point values.

As an additional attestation of management improvement, the interventions decreased worker turnover, likely reflecting increased management quality (Hoffman and Tadelis 2021) and thus head teacher and teacher satisfaction with their current postings. Teacher Training + Management Observation + People Skills reduced head teacher turnover by 12 percentage points (columns 3). The Teacher Training + Management Observation intervention increased the retention of teachers by 21 percentage points, increasing teacher retention by almost 50 percent relative to a control group mean of 49 percent (column 4). Teachers were 9 percentage points more likely to remain at their schools in the Teacher Training + Management Observation + People Skills intervention. For both retention measures we reject that the coefficients across the interventions were statistically equivalent—the additional People Skills training led to more head teachers remaining than the other arms but fewer teachers than the intervention that did not include People Skills. During the intervention year, teachers in both interventions indicated that head teachers and circuit supervisors were more likely to provide useful feedback, offer suggestions for improvement, and mention something that the

[37] GES instructed schools to continue with DI, inviting one teacher per school to a brief refresher training. No additional materials were provided.

[38] Effects on the individual components of the indices appear in Appendix Tables A18 and A19.

teacher did well, yet teachers in the People Skills Intervention also reported higher levels of job related stress and burnout. These countervailing forces likely led to the increased teacher retention in the People Skills intervention relative to the control group, but lower retention than in the non-People Skills Intervention. Both head teacher and teacher turnover could have reduced the persistent effects on management practices.

[Table 5 about here]

Teaching and Classrooms

Schools also continued to implement Differentiated Instruction the year after the intervention, demonstrating a permanent shift in teachers' beliefs about the value of Differentiated Instruction. About 42 percent of schools were still dividing their students by learning levels for one of the first two periods of the day (Table 6, column 1). While lower than during the intervention year, this level of implementation is about 7 times higher than was observed previously in Ghana and ten times higher than in India during the intervention years when management was not engaged (Duflo et al., 2024; Banerjee et al., 2016). Based on the same implementation index as in the intervention year, the average school was still implementing about 75 percent of the aspects of the DI program (column 2). The effect on the active learning index is still positive, but no longer statistically significant. Schools continued to divide students by learning level but did not maintain their active learning practices.

[Table 6 about here]

We test for heterogeneity by the degree of teacher turnover and whether the head teacher was still at the school. Schools with a higher proportion of retained teachers and those with a retained head teacher were more likely to be implementing the program (Appendix Table A20).[39]

Student Achievement

[39]Teacher retention and head teacher turn over were not random and could reflect other school characteristics that might lead to differential DI implementation.

Test scores of students who had been in treatment schools during the intervention year were still higher in July 2021, two full years after the end of the intervention (Table 7). Recall that in the year after the intervention, schools abruptly closed in March 2020 and remained closed until January 2021, reopening with a new school year. Further these students transitioned from primary school (P5 and P6) to junior high school (JHS1 and JHS2). Despite of all of this upheaval, the interventions increased student learning two years after the intervention by 0.06SD for students who had been in the Teacher Training + Management Observation schools and 0.08SD for students who had been in the Teacher Training + Management Observation + People Skills schools (column 1). Treatment students learned about 10 percent more than the control group relative to baseline and about 20 percent more than the control group learned between the first and second follow-up.[40] Figure 3 plots the short run and longer term test score trajectories of the three arms. Persistent test score increases after an intervention ends and students have changed schools are rarely observed.

[Table 7 about here]

[Figure 3 about here]

As with the results at the conclusion of the intervention, we do not find any heterogeneity by baseline test score, student gender, class at baseline, having literate parents, or at schools with better amenities (Appendix Table A21). For non-cognitive outcomes, students are equally likely to be enrolled in school and want to attend SHS, the next schooling level. Students from the Teacher Training + Management Observation schools were more likely than the other intervention and the control group to say that their favorite subject was math or English, a new difference relative to the first follow-up, and more likely than the control group to report liking math or English very much (Appendix Table A22), as was the case in the first follow-up. We do not find differential selection into test taking by the interaction

[40]These test score gains true gains and not the result of control school students' test scores decreasing—control school students increased their test score by 0.3SD between the two follow-up rounds.

of treatment status and baseline test score, but provide Lee (2008) bounds nevertheless (Appendix Tables A23 and A24).

The intervention improved the learning trajectories of students and buffered them against potential learning losses from school closures and transitions to higher levels of schooling.

7 Cost Effectiveness

We use the ingredients method to assess cost effectiveness (Hirji et al., 2023). Teacher Training + Management Observation cost $48 per student and the Teacher Training + Management Observation + People Skills cost $84 per student at the implemented scale in USD2018. As the achievement effect sizes were almost exactly equal across the two interventions, the Teacher Training + Management Observation intervention was clearly more cost effective. To scale this relative to the effect size per $100, this would be 0.23SD per $100 for Teacher Training + Management Observation and 0.13SD per $100 for Teacher Training + Management Observation + People Skills. To scale this as a price per year of learning in this context, Teacher Training + Management Observation costs $147 per student per year of learning and Teacher Training + Management Observation + People Skills costs $259 per student per year of learning.[41] The achievement effects persisted for at least two years and new cohorts of students continued to be treated with minimal additional costs, increasing the overall cost effectiveness.

8 Discussion

8.1 Conceptual Framework

In a public sector system with weak incentives, facilitating implementation is a non-trivial feat, yet reforms only increase productivity if they are implemented with fidelity (Banerjee et

[41] Both of these scaling methods make assumptions about the linearity of returns. We provide them to allow readers to compare this study's cost effectiveness to other studies.

al. 2017; Duflo et al., 2024; Angrist and Meager 2023). Our findings show that with existing systems and personnel, head teachers and circuit supervisors can be the key to encourage teachers to implement reforms.

This section lays out a conceptual framework based on the theory of leadership (Hermalin 1998; Bolton et al., 2012; Boudreau et al., 2021) to show how managers (head teachers) induced greater worker (teacher) effort by acting as leaders. As in the typical leader framework, head teachers are similar to workers (almost all were former teachers), but with additional experience, a higher wage, and a higher location in the bureaucratic hierarchy, making them more connected to senior management and better aware of the organizational priorities and agenda. In the conceptual framework, we account for this by assuming the head teacher has better information than the teachers and chooses his effort first.

Formally, consider a school with a single head teacher (HT) who manages N teachers indexed by $n = 1, \ldots, N$. We normalize achievement under the typical curriculum to be 0. School wide achievement beyond this level is a function of all teachers' DI efforts, $e_1, ..., e_n$, and the head teacher's effort, e_{HT}, times a productivity factor θ, plus an idiosyncratic error, ε. Formally,

$$A = \theta\, f(e_1, ..., e_n, e_{HT}) + \varepsilon \tag{2}$$

where $-1 \leq \theta \leq 1$ and θ is not known to teachers. This achievement includes both student achievement as measured by test scores and broader organizational priorities such as the school's overall performance and adherence to stated policies.

Teachers have beliefs about θ rooted in prior experiences. DI is new to these teachers–it was not how they were taught nor how they were taught to teach. Further, teachers might be uncertain about whether it is an organizational priority. Let $\tilde{\theta}$ reflect teachers' beliefs about θ, $-1 \leq \tilde{\theta} \leq 1$. Their skepticism about new teaching methods and how much the system values the intervention will cause their beliefs to understate the true return to DI, therefore $\tilde{\theta} \leq \theta$.

Teachers get utility from their fixed wages, achievement (both student and school), and school amenities and disutility from exerting effort. Each teacher acts in his own self interest in selecting his effort level in implementing DI, e, maximizing the following utility function

$$\max_{e} U\left(e_n\right) = w + A(\tilde{\theta}, e_1, ..., e_n, e_{HT}) - d\left(e_n\right) + F\left(V_s\right) \tag{3}$$

where w is the teacher's fixed wage, $A(.)$ is the achievement function from above but with a teacher's belief about the productivity and value of DI, $\tilde{\theta}$, instead of the true θ. V_s are a vector of school s amenities and attributes, and $d(e)$ is the disutility of DI effort, which is increasing in the amount of effort exerted.[42] The disutility function can be seen as a teacher's forgone utility from reducing his efforts spent on teaching the standard curriculum. The disutility from effort is an increasing and convex function with $d(0) = 0$. Teachers will select their e based on $\tilde{\theta}$, the expected return to DI, and resulting expected achievement $A = \tilde{\theta}\, f(e_1, ..., e_n, e_{HT}) + \varepsilon$. Underestimating θ induces the distortion in teacher effort—they put forth less effort on DI because they believe the return to their effort, $\tilde{\theta}$, is lower than the true return, θ. Optimal effort level is increasing in a teacher's belief about θ.

The head teacher at school s selects his own effort, e_{HT}, to maximize his utility as follows:

$$\max_{e_{HT}} U_{HT}\left(e_{HT}\right) = w_{HT} + \gamma A\left(.\right) - d\left(e_{HT}\right) + F\left(V_s\right)$$

where he gets utility from his set wage, w_{HT}, the achievement of the students and school $A\left(.\right)$ as specified above, a scaling parameter γ and other school amenities, $F\left(V_s\right)$.[43] He gets

[42] As with many civil servants, teachers' wages, w, do not depend on productivity or effort, but are instead set centrally based on years of experience and education. $A(.)$ enters a teachers' utility function to capture the ways in which student and school performance affects teacher utility without affecting their wages, for example because of intrinsic motivation, students are more pleasant to teach when they are learning, it could affect future classroom or school assignments, or they could be chided by parents, school officials, or other teachers for poor student performance or not adhering to institutional priorities. Both A and e may eventually affect w if they lead to a teacher becoming a head teacher, but selection of head teachers is mostly related to years of experience and education not student performance.

[43] As with teachers, head teachers' wages do not depend on student and school achievement but are set centrally and related to years of experience and education. $A(.)$ in the HT utility function captures the ways in which student and school performance affect head teacher utility without affecting their wages, for example because of intrinsic motivation, students being easier to manage when they are learning, it could

disutility from exerting DI effort, $d(e_{HT})$.

Credible signals about the magnitude of θ can increase $\tilde{\theta}$ and thus a teacher's effort. HTs can announce θ to teachers, for instance by saying that this new teaching method is effective or valued by the system. However, as the HT's payoff is increasing in the teachers' efforts, he has an incentive to announce the highest θ possible, something teachers realize, thus disregarding such announcements. Instead, if the HTs lead by spending their own time as a signal of value, they can induce teachers to update their own beliefs about θ (Hermalin 1998; Spence 1973).

Following Hermalin (1998), HTs choose their effort level first. Teachers observe this effort level, which informs teachers' own beliefs about θ. Specifically, HTs learn the true θ and choose a high e_{HT} because they realize the returns to effort are high.[44] Any HT effort that is unrelated to the achievement production function does not affect teacher belief about θ and does not increase teacher effort in DI.

The key to improving teacher effort is their $\tilde{\theta}$–their belief about the return to their own effort. When managers act as leaders they can change this belief. Changes to management practices that do not affect $\tilde{\theta}$ will not increase teacher effort. These changes could augment school amenities, V_s, and thus teacher utility, but will not feed into teacher effort.

Our interventions included visible, costly managerial effort focused on the return to DI—attending the training with the teachers and an observation form that provided a simple way for HT to demonstrate their increased effort in highlighting the most important aspects of DI implementation, i.e., a costly effort signal focused on DI. To complete the observation form, HT spent at least 30 minutes in a teachers' classroom, noting what was done well and areas for improvement. The first item of the observation form was, "Are learners grouped by learning level?" signaling that this was a priority undertaking and HTs were willing to sit in

affect future school assignments, or they could be chided by parents or school officials for poor student performance or not adhering to institutional priorities. Both A and e may eventually affect w if they lead to a head teacher promotion to district leadership.

[44] As managers, HTs have better information and more accurate beliefs about θ. In practice, this could be because they are higher up in the hierarchy and thus know that effort on DI and/or increased student achievement are institutional priorities on which they, their schools, and their teachers might be judged in the future. Higher level administrators could also exert costly effort to signal the true θ to HTs.

classrooms to ensure that it occurred. This signaled the size of θ to the teachers in a credible way since it was costly for the HT. Teachers then would exert more effort than they would under lower beliefs about θ. Teachers did update their beliefs about the return to DI and adjusted their effort accordingly (Table 2). Additional managerial practices and effort, like those conveyed in the People Skills training, only increase student learning if they affect the amount of teacher effort on DI. Any intervention that improves V_s increases teacher utility but does not affect teacher effort on DI.

8.2 Additional Discussion

Our results support the insight from our conceptual framework that when public sector managers act as leaders they can encourage reform in the absence of the private sector tools of management that compel workers to implement reforms, providing insight into how to scale effective interventions. In our setting, this leadership had a high return to school productivity by inducing increased teacher effort on effective classroom practices. Teachers reported more positive interactions with their managers when mangers had additional People Skills training, but as this job satisfaction did not result in changes to classroom practices, it did not improve student learning.

These findings rectify existing seemingly contradictory findings when comparing the observational and experimental literature on public sector management—not all management improvements are equivalent. The observational literature demonstrates the strong correlation between management quality and student test scores. In contrast, interventions that have improved school management have not improved test scores (Ganimian and Freel 2020; Muralidharan and Singh 2021; Romero et al. 2022). Our intervention with People Skills differentially increased management, but did not increase test scores beyond the Teacher Training + Management Observation intervention because what was happening in the classrooms was the same across the two interventions. To improve student learning, what is happening in the classroom has to change. Managers acting as leaders resulted in that

necessary classroom change.

Relatedly, our results provide empirical evidence on the "black box" of what types of school management matter for student learning. The statistically equivalent and almost identical point values in learning increases between the two treatment arms is consistent with Observational Practices being more important for student learning than People Skills or Other Management. In other words, parent meetings, school meetings, CS visits (independent of classroom observations), and whether teachers consider their managers good mentors are less important in the education production function as they do not change teacher effort or classroom practices. The simple rule of thumb observation form and encouragement to use it caused head teachers to focus more on instructional improvement and the additional People Skills training may have distracted them from exerting their leadership to change classroom practices. We found lower levels of active learning in the schools with People Skills relative to the treatment schools without it.

To further explore the relationship between Observational Practices, People Skills, and test scores, we conduct a mediation analysis (Bennet et al., 2018). Using 10,000 bootstrap samples, we pool the two treatments and estimate the effect of the receiving an intervention on the outcomes Observational Practices, People Skills, and test scores. We then correlate the treatment effects for the respective outcomes, testing whether in bootstrap samples with larger increases in a given measure of management there are also larger increases in test scores. Figure 4 presents the results of this exercise, demonstrating that there is a large, positive correlation between Observational Practices and test score improvements and a negative correlation between People Skills and test score improvements. These results support the interpretation that aspects of management that directly relate to pedagogy primarily explain our test score increases, while other aspects of management do not. Schools with larger gains in People Skills had the smaller test score increases as those with smaller gains in People Skills.

[Figure 4 about here]

The persistent effects one and two years after the end of the intervention show the lasting impact of managers acting as leaders to institutionalize reforms, promising results when considering scaling interventions. Teachers continued to implement the program, both Observational Practices and People Skills were higher than in the control schools, and both head teachers and teachers were more likely to stay in their existing schools. The Teacher Training + Management Observation intervention reduced teacher turnover by about 40 percent without any changes to incentive or pay–teachers were more satisfied to stay in their current schools potentially because their managers were acting as leaders through exerting their own visible costly effort, signaling that all members of the school were working to improve student achievement.

9 External Validity and Generalizability

Because public sectors in many settings are beset with the dual problems of insufficient provision of public services and weak managerial structures, our findings have many potential applications. In this section we formally assess the external validity and generalizability through an assessment of the selection, attrition, naturalness, and scale of the study.

Selection. Many countries, regardless of their levels of income, have at least some locations where students are behind grade level (List 2022). Similarly, many countries have public sectors with weak incentives. Our broad lessons are relevant to all of those countries. Our exact intervention might be most relevant to locations with similar characteristics, specifically the 53 countries with the same World Bank Lower-Middle Income categorization as Ghana (3 billion people), some of the richer countries in the 26 country Low Income category, some of the poorer countries in the 54 country Upper-Middle Income category, and the 47 countries of the sub-Saharan Africa region (1.2 billion people). The broad relevance is in part because our twenty districts are largely representative of Ghana but skew slightly poorer, coming from eight of the ten regions at the time, omitting Greater Accra and Brong-Ahafo,

two of the three most developed regions in Ghana in 2016 (UNICEF 2016).[45] The structure of teaching mimics other civil servant structures–teachers are hired and paid centrally with minimal control by local managers over terms or continuation of employment. Issues of how to motivate reticent civil servants in a minimal incentive environment are salient regardless of the country's level of income. Relative to other governments, Ghana's Worldwide Governance Indicators rating is above the average for sub-Saharan Africa, but below that of India's and the worldwide median (World Bank 2018). Therefore, Ghana's capacity does not stand out as unique but is similar to many countries. Further, students being in school but well behind grade level is not uniquely Ghanaian or even unique to middle income countries—it afflicts communities in some of the richest countries in the world.

Attrition. Attrition was very low in our study. None of our study schools left the sample during the course of the study. Student attrition was minimal, with 96 percent of students participating in the achievement follow-up. More details appeared in Section 6.3.

Naturalness. All tasks were ones that the participants could be expected to do as a part of their regularly assigned duties. When asked to implement a new teaching methodology or topic, teachers typically attend in-service teacher training events. Head teachers and circuit supervisors were already supposed to be observing teachers, this program provided them a more structured way in which to do so. As with teachers, when they are asked to implement new programs, they attend in-service training events.

Scale of the Study. The study mimicked how a similar program would be replicated at scale.The government designed, produced, and distributed all materials. All trainers were government employees. The teachers, head teachers, and circuit supervisor were existing government personnel completing tasks as part of their typical work. Since 2022, the government has been replicating this program at scale in 10,000 government schools.

Therefore, given the strong applicability of the issues in our sample, minimal attrition, and the high degree of both naturalness and implementation relative to how it would happen

[45]Our districts were between the 77th and 1st (poorest) percentile according to the national district poverty ranking in 2016 with an median percentile of 33.

at scale, this is a highly generalizable study both within education and the public sector more broadly.

10 Conclusions

Despite a recognition in the last 20 years of the importance of management for private sector productivity, less work has been done on the importance of management for public sector service delivery. We confirmed that management quality in Ghanaian schools is associated with higher student test scores, as has been previously tested in only seven other countries. Further, we implemented a 210 school, 3 arm randomized controlled trial in partnership with the Ministry of Education of Ghana to test the theory that school mangers, endowed with limited employment management tools, could instead act as leaders to encourage reform. We find that a costly management effort signal increased the likelihood that teachers implemented a new pedagogy, Differentiated Instruction, improving learning the same amount as a second intervention that also included People Skills training. The two interventions increased student test scores by 0.11 SD, about one third of a year of learning in this context. Despite only minimal training of one teacher per school the year after the end of the main intervention, we still observed compliance with Differentiated Instruction at a level that was 70 percent of the level during the initial supported year. Two years after the intervention, test score improvements persisted.

Getting existing teachers in existing systems to implement programs that have been shown to increase student learning is key to increasing student learning at scale. We show the importance of managers sending a costly effort signal, acting as leaders, to increase compliance with reforms that increased student learning.

References

Acemoglu, Daron and Matthew O. Jackson, "History, Expectations, and Leadership in the Evolution of Social Norms," *The Review of Economic Studies*, 2015, *82* (2 (291)), 423–456.

Akerlof, Robert and Richard Holden, "MOVERS AND SHAKERS," *The Quarterly Journal of Economics*, 2016, *131* (4), 1849–1874.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Andreoni, James, "Leadership giving in charitable fund-raising," *Journal of Public Economic Theory*, 2006, *8* (1), 1–22.

Angrist, Noam and Rachael Meager, "Implementation Matters: Generalizing Treatment Effects in Education," Technical Report, Working Paper 2023.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton, "From proof of concept to scalable policies: challenges and solutions, with an application," *Journal of Economic Perspectives*, 2017, *31* (4), 73–102.

_ , Shawn Cole, Esther Duflo, and Leigh Linden, "REMEDYING EDUCATION: EVIDENCE FROM TWO RANDOMIZED EXPERIMENTS IN INDIA," *The Quarterly Journal of Economics*, 2007.

Bank, World, *World Development Report: Learning to Realize Education's Promise. 2018*, World Bank, 2017.

Beg, Sabrin, Anne Fitzpatrick, and Adrienne M. Lucas, "Gender Bias in Assessments of Teacher Performance," *AEA Papers and Proceedings*, May 2021, *111*, 190–95.

Bennett, Daniel, Asjad Naqvi, and Wolf Peter Schmidt, "Learning, Hygiene and Traditional Medicine," *The Economic Journal*, 05 2018, *128* (612), F545–F574.

Bloom, Nicholas and John Van Reenen, "Measuring and explaining management practices across firms and countries," *The Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

_ and _ , "Human resource management and productivity," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1697–1767.

_ , Aprajit Mahajan, David McKenzie, and John Roberts, "Do management interventions last? Evidence from India," *American Economic Journal: Applied Economics*, 2020, *12* (2), 198–219.

_ , **Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

_ , **Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen**, "What Drives Differences in Management Practices?," *American Economic Review*, May 2019, *109* (5), 1648–83.

_ , **Renata Lemos, Raffaella Sadun, and John Van Reenen**, "Does management matter in schools?," *The Economic Journal*, 2015, *125* (584), 647–674.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ngángá, and Justin Sandefur, "Experimental evidence on scaling up education reforms in Kenya," *Journal of Public Economics*, 2018, *168*, 1 – 20.

Bolton, Patrick, Markus K. Brunnermeier, and Laura Veldkamp, "Leadership, Coordination, and Corporate Culture," *The Review of Economic Studies*, 12 2012, *80* (2), 512–537.

Boudreau, Laura, Rocco Macchiavello, Virginia Minni, and Mari Tanaka, "Union Leaders: Experimental Evidence From Myanmar," 2021. Working paper.

Brandts, Jordi, Christina Rott, and Carles Solà, "Not just like starting over-Leadership and revivification of cooperation in groups," *Experimental Economics*, 2016, *19*, 792–818.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico," *Journal of Political Economy*, 2018, *126* (2), 635–687.

Choi, Hye-Yeon, Eun Hye Kim, Joonsang Yoo, Kijeong Lee, Dongbeom Song, Young Dae Kim, Han-Jin Cho, Hyo Suk Nam, Kyung Yul Lee, Hye Sun Lee et al., "Decision-making support using a standardized script and visual decision aid to reduce door-to-needle time in stroke," *Journal of Stroke*, 2016, *18* (2), 239.

Cilliers, Jacobus, Brahm Fleisch, Cas Prinsloo, and Stephen Taylor, "How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching," *Journal of Human Resources*, 2020, *55* (3), 926–962.

_ , **Ibrahim Kasirye, Clare Leaver, Pieter Serneels, and Andrew Zeitlin**, "Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools," *Journal of Public Economics*, 2018, *167*, 69–90.

Coch, Lester and John RP French Jr, "Overcoming resistance to change," *Human relations*, 1948, *1* (4), 512–532.

Crawfurd, Lee, "School Management and Public Private Partnerships in Uganda," *Journal of African Economies*, 08 2017, *26* (5), 539–560.

Das, Jishnu, Stefan Dercon, James Habyarimana, and Pramila Krishnan, "Teacher shocks and student learning evidence from Zambia," *Journal of Human resources*, 2007, *42* (4), 820–862.

de Hoyos, Rafael, Alejandro J Ganimian, and Peter A Holland, "Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity Building for Public Schools in Argentina," *The World Bank Economic Review*, 11 2019, *35* (2), 499–520.

de Villiers, Rouxelle, "7 Principles of highly effective managerial feedback: Theory and practice in managerial development interventions," *The International Journal of Management Education*, 2013, *11* (2), 66–74.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas, "Experimental Evidence on Alternative Policies to Increase Learning at Scale," Working Paper 27298, National Bureau of Economic Research March 2021.

_ , _ , and Adrienne M Lucas, "Experimental Evidence on Four Policies to Increase Learning at Scale," *The Economic Journal*, 02 2024, *134* (661), 1985–2008.

Duflo, Esther, Rema Hanna, and Stephen P Ryan, "Incentives work: Getting teachers to come to school," *American Economic Review*, 2012, *102* (4), 1241–78.

Evans, David K. and Amina Mendez Acosta, "Education in Africa: What Are We Learning," Working Paper 542, Center for Global Development August 2020.

Fenizia, Alessandra, "Managers and Productivity in the Public Sector," *Econometrica*, 2022, *90* (3), 1063–1084.

Finan, Frederico, Benjamin A Olken, and Rohini Pande, "The personnel economics of the developing state," in "Handbook of Economic Field Experiments," Vol. 2, Elsevier, 2017, pp. 467–514.

Fitzpatrick, Anne, Sabrin Beg, Laura Derksen, Anne Karing, Jason Kerwin, Adrienne M. Lucas, Natalia Ordaz Reynoso, and Munir Squires, "Health knowledge and non-pharmaceutical interventions during the Covid-19 pandemic in Africa," *Journal of Economic Behavior and Organization*, 2021, *190*, 33–53.

G, Jr Fryer Roland, "Management and Student Achievement: Evidence from a Randomized Field Experiment," Working Paper 23437, National Bureau of Economic Research May 2017.

Ganimian, A. J. and S.H. Freel, "Can principal training improve school management? Short-term experimental evidence from Argentina," *Papeles de Economia Espanola*, 2021, *166*, 67–83.

Gershenson, Seth, "Performance standards and employee effort: Evidence from teacher absences," *Journal of Policy Analysis and Management*, 2016, *35* (3), 615–638.

Gibbons, Robert and John Roberts, "Economic Theories of Incentives in Organizations," in Robert Gibbons and John Roberts, eds., *The Handbook of Organizational Economics*, Princeton: Princeton University Press, 2013, pp. 56–99.

_ **and Rebecca Henderson**, "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises," in ROBERT GIBBONS and JOHN ROBERTS, eds., *The Handbook of Organizational Economics*, Princeton University Press, 2012, pp. 680–731.

Glewwe, Paul and Karthik Muralidharan, "Improving school education outcomes in developing countries: Evidence, Knowledge gaps, and Policy Implications," in "Handbook of Economics of Education," Vol. 5, North Holland, 2016, pp. 653–743.

Gray-Lobe, Guthrie, Anthony Keats, Michael Kremer, Isaac Mbiti, and Owen W Ozier, "Can Education be Standardized? Evidence from Kenya," *Evidence from Kenya (June 5, 2022). University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2022, (2022-68).

Haynes, Alex B, Thomas G Weiser, William R Berry, Stuart R Lipsitz, Abdel-Hadi S Breizat, E Patchen Dellinger, Teodoro Herbosa, Sudhir Joseph, Pascience L Kibatala, Marie Carmela M Lapitan et al., "A surgical safety checklist to reduce morbidity and mortality in a global population," *New England journal of medicine*, 2009, *360* (5), 491–499.

Hermalin, Benjamin E, "Toward an economic theory of leadership: Leading by example," *American Economic Review*, 1998, *88* (5), 1188–1206.

Herrmann, Mariesa A and Jonah E Rockoff, "Worker absence and productivity: Evidence from teaching," *Journal of Labor Economics*, 2012, *30* (4), 749–782.

Hirji, Shahana, Bethany Park, Edward Tsinigo, Sabrin Beg, Anne Fitzpatrick, and Adrienne Lucas, "Facilitating real-time cost collection and evaluating cost-effectiveness in a multi-armed study with government partners in Ghana," *Journal of Development Effectiveness*, 2023, *15* (1), 31–42.

Hoffman, Mitchell and Steven Tadelis, "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," *Journal of Political Economy*, 2021, *129* (1), 243–285.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Koukoumelis, Anastasios, M Vittoria Levati, and Johannes Weisser, "Leading by words: A voluntary contribution experiment with one-way communication," *Journal of Economic Behavior & Organization*, 2012, *81* (2), 379–390.

Lemos, Renata and Daniela Scur, "Developing management: An expanded evaluation tool for developing countries," *London School of Economics, Centre for Economic Performance, London*, 2016.

_ , **Karthik Muralidharan, and Daniela Scur**, "Personnel management and school productivity: Evidence from india," Technical Report, National Bureau of Economic Research 2021.

List, John A., *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*, New York: Random House, 2022.

List, John A and Daniel Rondeau, "The impact of challenge gifts on charitable giving: an experimental investigation," *Economics Letters*, 2003, *79* (2), 153–159.

_ **and David Lucking-Reiley**, "The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign," *Journal of political Economy*, 2002, *110* (1), 215–233.

Loeper, Antoine, Jakub Steiner, and Colin Stewart, "Influential Opinion Leaders," *The Economic Journal*, 2014, *124* (581), 1147–1167.

Lucas, Adrienne M., "Selection, training, and importance of school heads and supervisors across Africa," *International Journal of Educational Development*, 2024, *109*, 103094.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian, "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 2019, *109* (4), 1426–1460.

_ **and** _ , "Improving Public Sector Management at Scale? Experimental Evidence on School Governance India," Working Paper 28129, National Bureau of Economic Research November 2020.

_ **and Paul Niehaus**, "Experimentation at scale," *Journal of Economic Perspectives*, 2017, *31* (4), 103–24.

_ , _ , **and Sandip Sukhtankar**, "Building state capacity: Evidence from biometric smartcards in India," *American Economic Review*, 2016, *106* (10), 2895–2929.

Oliver, Regina M, Joseph H Wehby, and J Ron Nelson, "Helping teachers maintain classroom management practices using a self-monitoring checklist," *Teaching and Teacher Education*, 2015, *51*, 113–120.

Potters, Jan, Martin Sefton, and Lise Vesterlund, "Leading-by-example and signaling in voluntary contribution games: an experimental study," *Economic Theory*, 2007, *33*, 169–182.

Rasul, Imran and Daniel Rogger, "Management of bureaucrats and public service delivery: Evidence from the nigerian civil service," *The Economic Journal*, 2018, *128* (608), 413–446.

Romero, Mauricio, Juan Bedoya, Monica Yanez-Pagans, Marcela Silveyra, and Rafael de Hoyos, "Direct vs indirect management training: Experimental evidence from schools in Mexico," *Journal of Development Economics*, 2022, *154*, 102779.

_ , Justin Sandefur, and Wayne Aaron Sandholtz, "Outsourcing education: Experimental evidence from Liberia," *American Economic Review*, 2020, *110* (2), 364–400.

Semel, Marcus E, Stephen Resch, Alex B Haynes, Luke M Funk, Angela Bader, William R Berry, Thomas G Weiser, and Atul A Gawande, "Adopting a surgical safety checklist could save money and improve the quality of care in US hospitals," *Health Affairs*, 2010, *29* (9), 1593–1599.

Spence, A. Michael, "Time and Communication in Economic and Social Interaction*," *The Quarterly Journal of Economics*, 11 1973, *87* (4), 651–660.

Tankard, Margaret E and Elizabeth Levy Paluck, "Norm perception as a vehicle for social change," *Social Issues and Policy Review*, 2016, *10* (1), 181–211.

Tsai, Thomas C, Ashish K Jha, Atul A Gawande, Robert S Huckman, Nicholas Bloom, and Raffaella Sadun, "Hospital board and management practices are strongly related to hospital performance on clinical quality metrics," *Health affairs*, 2015, *34* (8), 1304–1311.

Vivalt, Eva, "How much can we generalize from impact evaluations?," 2016.

A Appendix for Online Publication

A.1 Additional Intervention Details

Teacher Material Development

Since the success of a program at scale depends on government systems supporting and sustaining the intervention, this intervention involved Ghana Education Services (GES) and its subsidiaries including the Basic Education Division (BED), National Council for Curriculum and Assessment (NaCCA), and the National Inspectorate Board (NIB). All materials, personnel, and training were implemented through the existing government system except the UNICEF intern who monitored the barely used help-desk.

The National Council for Curriculum and Assessment (NaCCA), the board responsible for managing and implementing government curriculum policy, led the development of DI materials. A NaCCA Resource Development Team conducted an initial review of existing GES materials and materials used in the previous implementation of targeted instruction in Ghana, the Teacher Community Assistant Initiative (TCAI). They then modified these existing resources and designed and developed new materials as necessary. The teaching materials included topics to cover each day, and ideas for class, group, and individual activities. These were not scripted lessons, leaving teachers the latitude to pick the activities that most resonated with their students.

A core team of the National Teaching Council (NTC), the council responsible for licensing teachers in Ghana, developed the training materials and facilitator manuals.

Teacher Training

Teacher training was a cascade model. NTC trained 24 national trainers. The national trainers trained 160 District Teacher Support Team (DTST) members, i.e. district-level government employees regularly responsible for in-service teacher training. These DTST members trained the treatment teachers, head teachers, and circuit supervisors. The training included time to practice Differentiated Instruction with their peers.

The main teacher training occurred for 5 days prior to the start of the 2018-2019 academic year with shorter refresher trainings for 3 days at the start of Term 2 and 2 days at the start of Term 3.

The DI materials and methods focused on strong foundational learning and full comprehension of concepts instead of rote memorization. At the start of each term, teachers were to assess all students in grades 4 through 6 to determine their learning levels. Teachers adjusted students' levels as necessary throughout the year. DI continued through terms 2 and 3 to the end of the academic year.

People Skills Material Development

The National Inspectorate Board (NIB), the board that oversees school inspection and evaluation, developed the circuit supervisors' and head teachers' People Skills intervention materials including the their manuals, quick reference cards, and content for text message reminders. Much of the content was a condensed version of the existing manual for head teachers and circuit supervisors, *Leadership for Learning.* The management manuals were not limited to DI content, instead reminding head teachers and circuit supervisors how to be effective school leaders and support teachers, broadly. The training focused on guidelines for productive interpersonal relationships, including mentoring and coaching.

People Skills Training

The People Skills training occurred after the teacher training and prior to the start of the 2018-2019 academic year. The initial training was 3 days with 2 days of refresher training prior to the start of Term 2, and a 1 day refresher training prior to the start of Term 3.

Implementation

The training recommended implementing differentiated instruction in one of the first two periods of the day, on Tuesdays through Fridays, and starting the third week of the term. About 43 percent of head teachers reported that their school implemented DI in Period 1, 35 percent in Period 2, and 22 percent at another time of day.

Students were placed in levels using a simple tool modeled after the Annual Status of

Education Report (ASER) assessment in India. This was not the same tool that was used to estimate the effect of the intervention on student learning.

In the first two terms the teaching observation forms were completed on paper. In term 3, these forms were completed on a tablet through the mobile School Report Card (mSRC) system that all head teachers, both treatment and control, had.

A.2 Test Construction

We designed the test to include both foundational concepts and grade level content. The majority of questions were adapted from the examinations used in Duflo, Kiessel, and Lucas (2024), which had been originally developed by education stakeholders in the Ministry of Education to reflect grade 1 through grade 4 material. We added additional questions that were inspired by questions from the Ghana National Education Assessment grade 3 and 6 exams. Enumerators conducted the assessments one-on-one. Based on piloting and data from Duflo, Kiessel, and Lucas (2024), the exam questions were divided into three difficulty levels: easier, medium, and harder. Enumerators conducted the assessments one-on-one with students. In consideration of student and enumerator time, not all students were asked all questions. The tests were semi-adaptive: all students started with the medium questions, then progressed to either the easier or harder questions based on their performance. As all students completed the same anchor (medium) questions, we used item response theory to put all scores on a common scale. Teachers did not see the exams nor did students retain any papers that they could have shared with their teachers. These were not the same exams that teachers used to level students. Appendix Figure A3 contains the test score distribution at baseline.

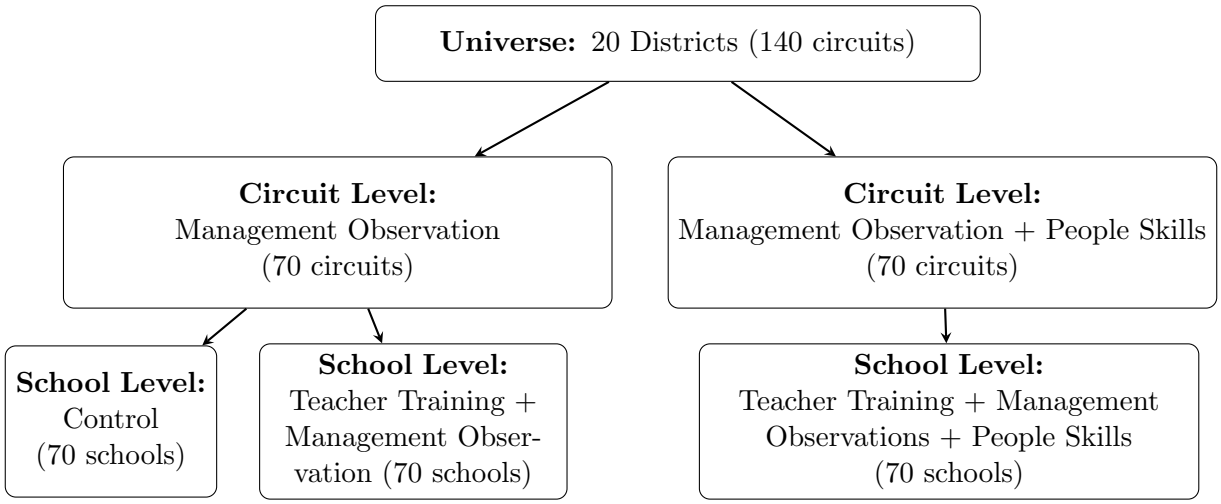
[Appendix Figure A3 about here]

A.3 Additional Tables

Tables A1 through A23 contain additional supporting materials.

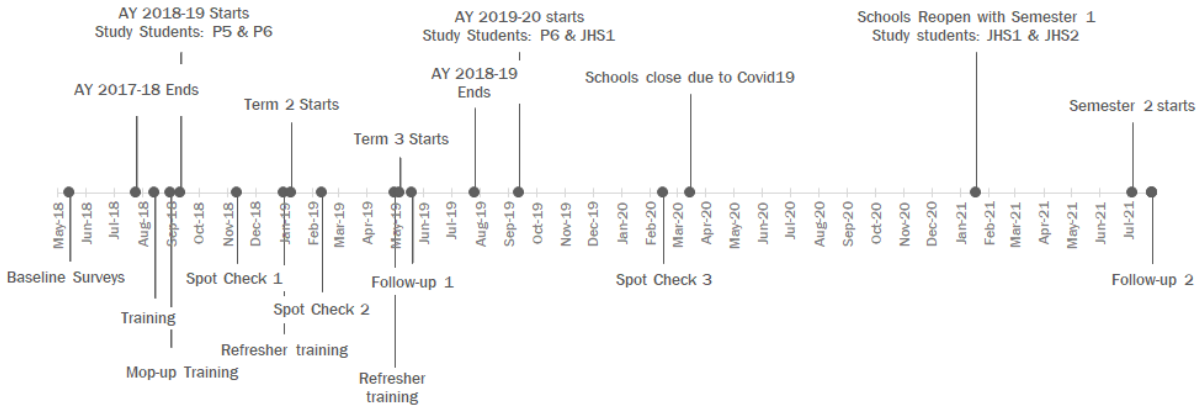
[Appendix Table A1-A24 about here]

Figure 1: Randomization Design



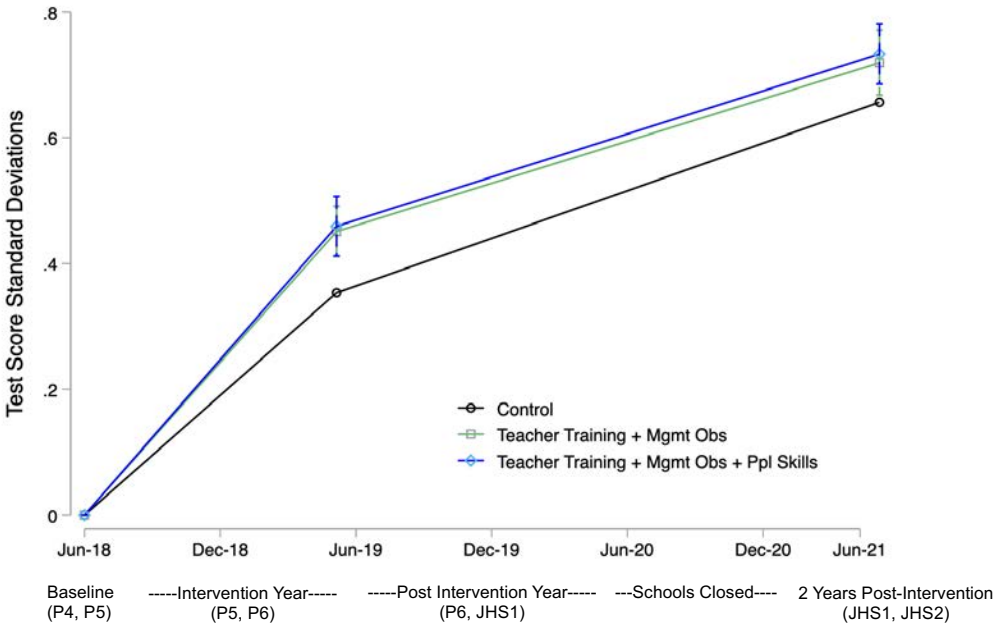
Notes: A circuit is a geographically proximate group of approximately 9 schools assigned to a single circuit supervisor. The study involved only 1 or 2 schools per circuit. Circuits were randomly assigned to the Management Observations arm or the Management Observations + People Skills treatment. Within each circuit, 2 eligible schools were randomly selected and randomized into control or Teaching Training + Management Observations in the Management Observations circuits and Teacher Training + Management Observations + People Skills or removed from the sample in the Management Observations + People Skills circuits.

Figure 2: Timeline of Study Activities



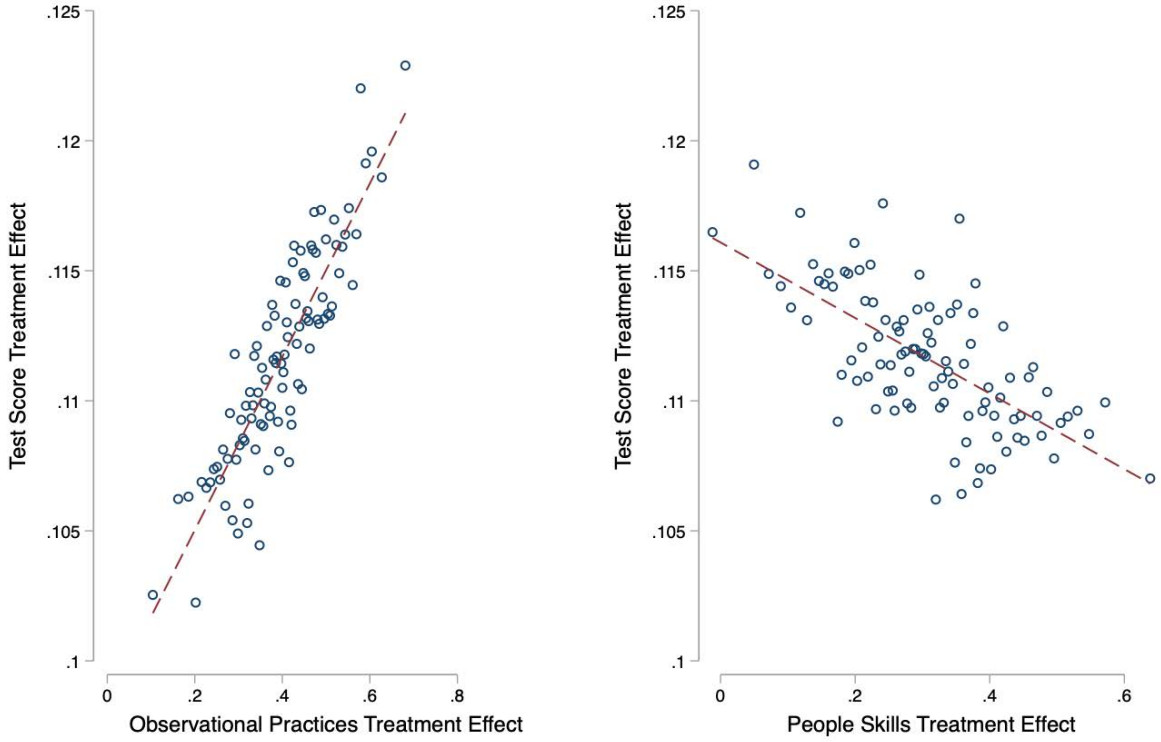
Notes: Activities above the timeline correspond to school calendar dates, while activities below the timeline correspond to intervention and data collection activities.

Figure 3: Short and Longer Term Test Score Effects



Notes: Above is the trajectory of learning gains over the course of the study by treatment group. Student assessments were conducted at baseline and two follow-ups. The error bars are 95% confidence intervals. The students' grade level on the x-axis assumes timely progression. Baseline students were testing regardless of grade level at follow-up.

Figure 4: Mediation Analysis: Correlation of Management Indices and Test Score Treatment Effects



Notes: Using 10,000 bootstrapped samples we estimated pooled treatment effects for test scores, observational practices, and people skills indices. Graphs are binned scatterplots of the estimates with the best-fit line included. The left graph shows the correlation between the treatment effects on test scores and the observational practices index. The right graph shows the correlation between the treatment effects on test scores and people skills.

Table 1: Effects on Management Practices

	Observational Practices Index (1)	**People Skills Index** (2)	**Other Management Index** (3)
Teacher Training + Mgmt Obs	0.397***	0.162	0.367**
	(0.133)	(0.141)	(0.144)
Teacher Training + Mgmt Obs + Ppl Skills	0.401***	0.468***	0.641***
	(0.125)	(0.151)	(0.165)
P-Value Same Effect	0.98	0.05	0.07
Observations	210	210	210
R^2	0.30	0.32	0.21
Mean Dep., Control	-0.00	0.00	0.00

Notes: Robust standard errors appear in parenthesis. Column 1: Linear probability model. Columns 2-4: The outcome variables are standardized indices of management outcomes. All regressions include baseline management index and district fixed effects.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 2: Effects on Classroom Activities

Dependent Variable:	**Teacher in Classroom** (1)	**Active Learning Index** (2)	**Class in Levels** (3)	**DI Imp Index** (4)
Teacher Training + Mgmt Obs	0.130***	0.470***	0.572***	0.787***
	(0.032)	(0.060)	(0.047)	(0.018)
Teacher Training + Mgmt Obs + Ppl Skills	0.111***	0.362***	0.608***	0.818***
	(0.034)	(0.059)	(0.046)	(0.017)
P-Value Same Effect	0.51	0.06	0.53	0.19
Observations	2,462	2,462	420	210
R^2	0.07	0.10	0.43	0.92
Mean Dep., Control	0.68	0.00	0.00	0.00

Notes: Regressions include and strata (district) fixed effects. Standard errors clustered at the school level appear in parenthesis. Columns 1 and 3: Linear probability models. Columns 1 and 2: Unit of observation is class period by spot check round. Column 1: Whether the teacher was present during the entire classroom observation; also includes a control for whether the observation is from the first or second spot check, whether the observation is from the first or second class period observed, and the average percent of teachers present during the baseline survey. Column 2: An index of active learning variables (teacher presence; any TLM use; teacher engaged; student expresses an idea). Column 3: Whether or not the students were observed to be split by levels instead of by class during the spot check. Unit of observation is school in each spot check round. Column 4: The percent of 5 different aspects of Differentiated Instruction that the respondent reported or was observed doing, averaged at the school level. Unit of observation is school-level.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3: Effects on Norms of Behavior

	Teacher Opinions of Teacher Shirking	**Teacher Opinions of All Roles Shirking**	**HT Opinions of Teacher Shirking**	**HT Opinions of All Roles Shirking**
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Obs	-0.261**	-0.310***	-0.251*	-0.038
	(0.108)	(0.105)	(0.152)	(0.153)
Teacher Training + Mgmt Obs + Ppl Skills	-0.354***	-0.401***	-0.301**	-0.206
	(0.107)	(0.118)	(0.142)	(0.156)
P-Value Same Effect	0.36	0.40	0.74	0.28
Observations	461	461	209	209
R^2	0.28	0.28	0.26	0.28
Mean Dep., Control	-0.00	0.00	-0.00	-0.00

Notes: Outcomes in all columns are indices of questions about hypothetical people in a particular role, grouped by respondent type, standardized by control group mean and standard deviation. Higher values indicate that shirking is more acceptable. Regressions include respondent age, gender, experience, age squared and experience squared, and district and survey round fixed effects. Standard errors clustered at the school level appear in parentheses. Column 1: Index of assessments of hypothetical teachers by teachers. Column 2: Index combining all teacher responses. Column 3: Index of assessments of hypothetical teachers by head teachers. Column 4: Index combining all head teacher responses.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 4: Effects on Student Test Scores

	Combined Score (1)	**Math Score** (2)	**English Score** (3)
Teacher Training + Mgmt Obs	0.108***	0.140***	0.065***
	(0.021)	(0.026)	(0.022)
Teacher Training + Mgmt Obs + Ppl Skills	0.107***	0.131***	0.076***
	(0.024)	(0.029)	(0.024)
P-Value Same Effect	0.95	0.75	0.63
Observations	5,608	5,608	5,608
R^2	0.74	0.63	0.71
Mean Dep., Control	0.33	0.32	0.30

Notes: Regressions include controls for student age, age-squared, grade at baseline, baseline assessment scores, female, and strata (district) fixed effects. Standard errors clustered at the school level appear in parenthesis. Scores standardized relative to the baseline pooled sample.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 5: Management Outcomes in Year After Intervention

	Observational Practices	**People Skills**	**HT Retained**	**Teacher Retained**
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Obs	0.146	0.451**	-0.010	0.206***
	(0.129)	(0.210)	(0.071)	(0.042)
Teacher Training + Mgmt Obs + Ppl Skills	0.262**	0.851***	0.117*	0.092**
	(0.129)	(0.227)	(0.071)	(0.046)
P-Value Same Effect	0.30	0.11	0.07	0.01
Observations	208	208	208	687
R^2	0.33	0.32	0.26	0.16
Mean Dep., Control	0.00	0.00	0.70	0.49

Notes: Based on data collected at the unannounced spot check in the year following the intervention. All columns include district fixed effects. Columns 1-2: Includes baseline management index. The outcome variables are standardized indices of management outcomes. Columns 3-4: linear probability models. Indicator of whether head teacher or teacher was still at the same school. Column 4: standard errors clustered at the school level.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 6: Effects on Classroom Activities in Year After Intervention

Dependent Var:	**Class In Levels**	**DI Imp Index**	**Active Learning Index**
	(1)	(2)	(3)
Teacher Training + Mgmt Obs	0.416***	0.723***	0.047
	(0.072)	(0.038)	(0.135)
Teacher Training + Mgmt Obs + Ppl Skills	0.448***	0.767***	0.202
	(0.091)	(0.044)	(0.165)
P-Value Same Effect	0.77	0.43	0.32
Observations	127	127	363
R^2	0.39	0.81	0.23
Mean Dep., Control	0.00	0.00	-0.00

Notes: Regressions include whether the observation is from the first or second class period observed, the average percent of teachers present during the baseline survey, and strata (district) fixed effects. Standard errors clustered at the school level appear in parenthesis. Column 1: Unit of observation is school. Whether the class was split by levels during the spot check. Linear probability model. Column 2: The portion of 5 different measures of differentiated instruction that the respondent reports doing. Unit of observation is school. Control group was imputed to be zero. Column 3: An index of active learning variables. Unit of observation is school-classroom observation.
$*p < 0.10, **p < 0.05, ***p < 0.01$

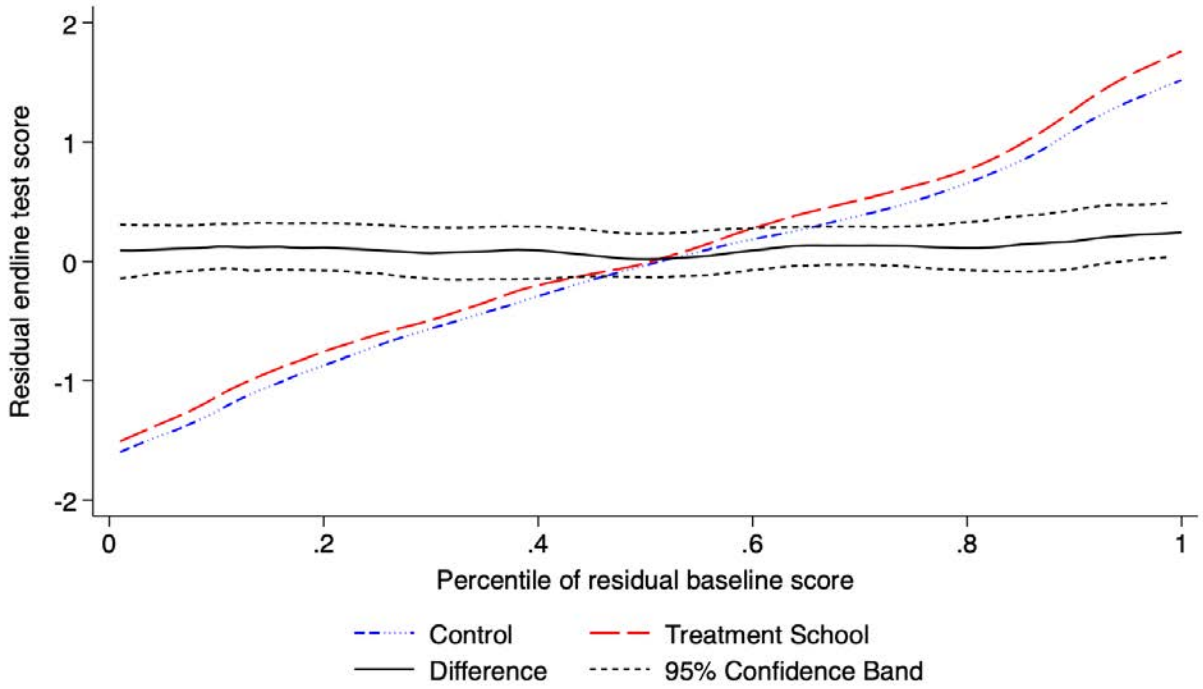
Table 7: Persistent Student Test Scores Effects

	Combined Score (1)	**Math Score** (2)	**English Score** (3)
Teacher Training + Mgmt Obs	0.062**	0.077**	0.055**
	(0.026)	(0.031)	(0.027)
Teacher Training + Mgmt Obs + Ppl Skills	0.076***	0.084***	0.073***
	(0.024)	(0.031)	(0.025)
P-Value Same Effect	0.62	0.82	0.55
Observations	5,080	5,080	5,080
R^2	0.66	0.55	0.63
Mean Dep., Control	0.66	0.61	0.63

Notes: Sample is all students available for the assessments at the second follow-up. Regressions include controls for student age, age-squared, grade at baseline, baseline assessment scores, female, and strata (district) fixed effects. Standard errors clustered at the school level appear in parenthesis. Scores standardized relative to the baseline pooled sample.
$*p < 0.10, **p < 0.05, ***p < 0.01$

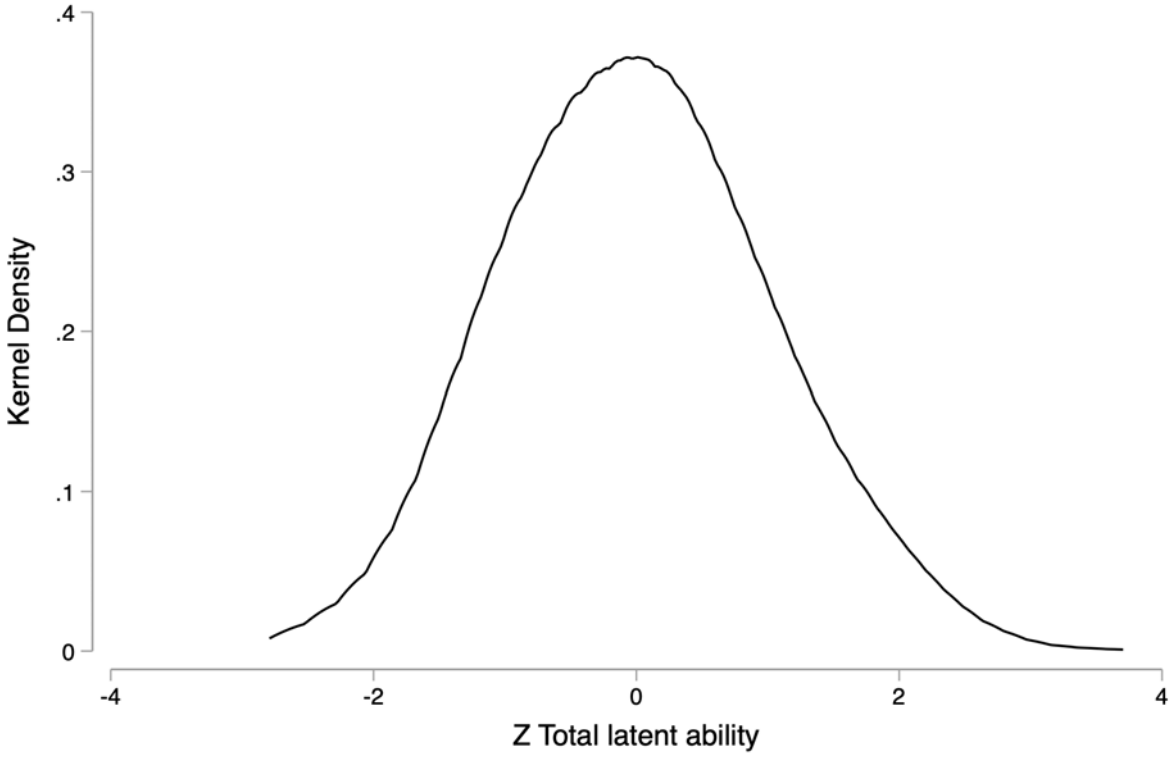
Appendix

Figure A1: Non-Parametric Test Results



Notes: This plot shows the non-parametric test score effects of either treatment on the combined test score. Test score effects are residulalized by baseline student age, age-squared, grade at baseline, baseline assessment scores, female, and strata (district) fixed effects. The red dashed line is the treatment group, and the blue dashed line the control group. The difference between the treatment and the control group is the black line. The short dashed lines display the 95 percent confidence intervals.

Figure A2: Distribution of Student Achievement at Baseline



Notes: Plot shows the standardized distribution of the combined IRT math and English score of all students at baseline.

Appendix Tables

Table A1: Components of Management Indices and Mapping of Indices to D-WMS

Panel A: Observational Practices	**WMS Category**
Panel A: Observational Practices	
Number CS classroom observations (HT)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
HT Classroom Observations (5 min) (T)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
Average Number of P4-P6 Classroom Observations (5 min) by HT (HT)	2.2: Implementation and Monitoring & 3.3: Implementation and Monitoring
CS Classroom Observations (5 min) (T)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
HT Classroom Observation Feedback (T)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
CS Classroom Observation Feedback (T)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
Average Number of P4-P6 Classroom Observations (30 min) by HT (HT)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
How often does HT observe teaching in school (HT)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
HT Classroom Observations (30 min) (T)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
CS Classroom Observations (30 min) (T)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
Number of CS Classroom Observations (5 min) in study schools (CS)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
Number of CS Classroom Observations (30 min) in study schools (CS)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
Number of school visits by CS (HT)	2.2: Implementation and Monitoring 3.3: Implementation and Monitoring
Panel B: People Management	
HT provides constructive feedback (HT)	6.1: Finding and documenting problems 6.2: Who resolves problems 9.3: Get people involved in constructive feedback
How often HT gives Ts suggestions about improving teaching (HT)	6.1: Finding and documenting problems 6.2: Who resolves problems /9.3: Get people involved in constructive feedback
An important part of HT job is to ensure teaching skills are improving (HT)	6.3: Who improves processes
HT encourages teachers to try new practices (HT)	5.2: How are these best practices shared within the school
How often HT takes initiative to discuss matters with Ts (HT)	6.1: Finding and documenting problems 6.2: Who resolves problems /9.3: Get people involved in constructive feedback
Number of staff meetings each term by HT (HT)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school
Number of HT meetings with CS (HT)	12.3: Goals are well communicated within the school
CS is a valuable mentor (HT)	6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
CS works with HT to solves problems (HT)	6.1: Finding and documenting problems 6.2: Who resolves problems 9.3: Get people involved in constructive feedback /10.2: How long it takes to identify and deal with a problem
CS performs valuable work for school (HT)	6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
CS is a good manager (HT)	6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
Number of staff meetings by HT (T)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school
Number of staff meetings by CS (T)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school
CS provides useful feedback (T)	9.3: Get people involved in constructive feedback
HT provides useful feedback HT (T)	9.3: Get people involved in constructive feedback
HT feedback mentions something T did well (T)	9.3: Get people involved in constructive feedback
HT feedback offers sugg. for improvement (T)	6.3: Who improves processes 9.3: Get people involved in constructive feedback
CS feedback offers sugg. for improvement (T)	6.3: Who improves processes 9.3: Get people involved in constructive feedback
CS feedback mentions something T did well (T)	9.3: Get people involved in constructive feedback
T feel valued and appreciated (T)	23.2: Teachers are aware of the benefits of working at your school
HT has helped T become a better Teacher (T)	9.3: Get people involved in constructive feedback
CS has helped T become a better Teacher (T)	9.3: Get people involved in constructive feedback
HT is a good manager (T)	6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
Number of CS meetings with each HT in their circuit (CS)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school
Number of CS meetings with all staff their circuit (CS)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school
Number of CS visits to schools their circuit (CS)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school
Number of CS meetings with all HTs their circuit (CS)	12.1: Clarity of goals through the hierarchy chain 12.3: Goals are well communicated within the school

Suggestions to teachers frequency (CS)	9.3: Get people involved in constructive feedback
Strongly agree CS valuable mentor for teachers (CS)	6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
Strongly agree CS valuable mentor for headteachers (CS)	6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
Strongly agree CS provides constructive feedback to teachers (CS)	9.3: Get people involved in constructive feedback
Strongly agree CS perform valuable work for school (CS)	6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
CS provides suggestions to HT (CS)	6.1: Finding and documenting problems 6.2: Who resolves problems 9.3: Get people involved in constructive feedback
Other Management	
An important part of HT job is to ensure teachers are held accountable (HT)	6.1: Finding and documenting problems 6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
HT scheduled meetings with parents (HT)	3.2: Student/parent engagement in student learning
Teachers receive rewards for good performance (HT)	18.1: Identification of good performers
School had student attendance records (HT)	4.1: Individual student data availability
School had teacher attendance records (HT)	7.2: Performance tracking frequency
Schools uses teacher feedback to guide goals (HT)	11.1: Clarity and Balance of Targets/Goal Metrics
Teachers share new practices with other teachers (HT)	5.2: How are these best practices shared within the school
Teachers encourage students to approach them for supplemental help (HT)	3.2: Student/parent engagement in student learning
The school uses student scores to establish teachers' effectiveness (HT)	4.3: School management of critical student transitions
HT presence at spot checks (Spot check)	15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
An important part of CS job is to ensure teaching skills are improving (CS)	6.2: Who resolves problems 6.3: Who improves processes
An important part of CS job is to ensure teachers are held accountable (CS)	6.1: Finding and documenting problems 6.2: Who resolves problems 6.3: Who improves processes 15.1: Who is responsible for school reaching overall targets (including quality, equity, and cost-effectiveness of student outcomes)
Schools have goals/school development plan (CS)	11.1: Clarity and Balance of Targets/Goal Metrics 13. Time Horizon of Targets/Goals
CS tried out new ideas in school (CS)	3.1: Identifying and addressing individual student needs
Schools use student scores to guide school goals (CS)	11.2: Goals set at the district, school, departmental and individual levels

Notes: Abbreviation in parenthesis indicates respondent: T=Teacher. HT=Head Teacher. CS=Circuit Supervisor.

Table A2: Summary Statistics

	—Control—	Tch Training+Mgmt Obs	Tch Training+Mgmt Obs+ Ppl Skills	P-Value of Equality
Panel A: Student-Level Variables	(1)	(2)	(3)	(4)
Male	0.53	0.54	0.53	0.83
	(0.50)	(0.50)	(0.50)	
P4 student	0.50	0.51	0.50	0.14
	(0.50)	(0.50)	(0.50)	
Age	12.17	12.05	12.12	0.62
	(1.77)	(1.88)	(1.83)	
Baseline Math	0.01	-0.02	0.01	0.84
	(0.99)	(0.99)	(1.02)	
Baseline English	-0.00	0.00	0.00	1.00
	(0.97)	(1.01)	(1.00)	
Baseline Composite Score	0.00	-0.01	0.01	0.97
	(0.98)	(1.00)	(1.02)	
N	2031	1932	1930	
Panel B: Teacher-Level Variables				
Male	0.74	0.75	0.73	0.94
	(0.44)	(0.44)	(0.46)	
Age	31.64	31.57	30.96	0.18
	(6.90)	(7.38)	(6.05)	
Teacher Has BA	0.27	0.23	0.21	0.34
	(0.44)	(0.42)	(0.41)	
Years Experience as a Teacher	6.17	6.13	5.95	0.63
	(5.97)	(6.10)	(5.44)	
Teacher Present at Arrival	0.84	0.86	0.89	0.68
	(0.36)	(0.35)	(0.31)	
# HT Class Obs (Terms 1 + 2, Any Length)	9.03	8.81	8.04	0.86
	(15.11)	(12.57)	(12.93)	
HT Gives Feedback About Teaching	0.74	0.77	0.79	0.56
	(0.44)	(0.42)	(0.41)	
N	217	226	228	
Panel C: Head Teacher and School-Level Variables				
Male	0.81	0.86	0.86	0.73
	(0.39)	(0.35)	(0.35)	
Age	42.90	40.67	42.84	0.19
	(8.84)	(8.18)	(9.54)	
Years Experience as HT	6.81	6.03	7.64	0.15
	(5.93)	(4.90)	(5.94)	
School Enrollment P4-P5	71.83	80.69	74.46	0.30
	(32.35)	(48.18)	(50.57)	
N	70	70	69	
Panel D: CS-Level Variables				
Male		0.92	0.89	0.43
	–	(0.27)	(0.31)	
Age		45.93	44.36	0.19
	–	(6.37)	(7.80)	
Number of Schools in Circuit	–	8.36	8.40	0.93
	–	(2.71)	(2.33)	
Years Experience as CS	–	3.72	3.52	0.55
		(3.09)	(3.18)	
N	–	70	70	

Notes: Each cell presents the mean, with standard deviations in parentheses, from the baseline survey. One head teacher declined the baseline survey. Column 4 presents the p-value on the F-test of joint equality for columns 1−3, controlling for strata (district) and using standard errors clustered at the school level. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A3: Effects on Individual Components of Management Observation Practices Index

	Teacher Training + Mgmt Obs (1)	**Teacher Training + Mgmt Obs + Ppl Skills** (2)	**P-value** (3)	**Control Mean** (4)
Average Number of P4-P6 Classroom Observations (5 min) by HT (HT)	0.265	-0.599	0.019	3.299
	(0.489)	(0.418)		
Number CS classroom observations (HT)	1.814***	2.862***	0.045	2.171
	(0.429)	(0.430)		
Number of school visits by CS (HT)	1.643**	2.224***	0.459	5.971
	(0.695)	(0.721)		
HT Classroom Observations (5 min) (T)	0.589	0.050	0.113	3.284
	(0.387)	(0.313)		
CS Classroom Observations (5 min) (T)	0.612***	0.706***	0.454	0.779
	(0.111)	(0.126)		
HT Classroom Observation Feedback (T)	0.173***	0.210***	0.265	0.591
	(0.037)	(0.036)		
CS Classroom Observation Feedback (T)	0.281***	0.316***	0.350	0.305
	(0.035)	(0.036)		
Average Number of P4-P6 Classroom Obsevations (30 min) by HT (HT)	0.936***	0.827***	0.580	0.849
	(0.191)	(0.201)		
How often does HT observe teaching in school (HT)	0.171	0.243*	0.575	3.371
	(0.132)	(0.129)		
HT Classroom Observations (30 min) (T)	0.707***	0.886***	0.239	0.725
	(0.144)	(0.123)		
CS Classroom Observations (30 min) (T)	0.456***	0.664***	0.021	0.206
	(0.072)	(0.082)		
Number of CS visits to study school (CS)	0.733**	2.024***	0.001	4.211
	(0.346)	(0.389)		
Number of CS classroom observations (5 min) in study school (CS)	0.856***	1.564***	0.005	1.958
	(0.220)	(0.226)		

Notes: Based on data collected at the first follow-up. Columns 1 and 2: Coefficient estimates from Equation (1) for the listed component of the Management Observation Practices index. All regressions include baseline management index and district fixed effects. Robust standard errors appear in parentheses. Column 1: Effect of Teacher Training + Management Observation. Column 2: Effect of the Teacher Training + Management Observation + People Skills intervention. Column 3: the p-value on the null hypothesis that the two effects are statistically equivalent. Column 4: the control group mean.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A4: Effects on Individual Components of People Skills Index

	Teacher Training + Mgmt Obs	Teacher Training + Mgmt Obs + Ppl Skills	P-value	Control Mean '
	(1)	(2)	(3)	(4)
HT provides constructive feedback (HT)	0.071	0.053	0.827	4.471
	(0.081)	(0.088)		
How often HT gives Ts suggestions about improving teaching (HT)	-0.029	-0.044	0.873	3.686
	(0.101)	(0.105)		
An important part of HT job is to ensure teaching skills are improving (HT)	0.114	0.132*	0.812	4.557
	(0.081)	(0.078)		
HT encourages teachers to try new practices (HT)	0.143*	0.113	0.719	4.314
	(0.079)	(0.083)		
How often HT takes initiative to discuss matters with Ts (HT)	0.059	-0.011	0.533	3.551
	(0.111)	(0.111)		
Number of staff meetings each term by HT (HT)	-0.156	0.277*	0.006	2.407
	(0.128)	(0.145)		
Number of HT meetings with CS (HT)	-0.067	0.894	0.160	4.217
	(0.571)	(0.576)		
CS is a valuable mentor (HT)	0.071	0.181*	0.290	4.386
	(0.097)	(0.103)		
CS works with HT to solves problems (HT)	0.086	0.153	0.479	4.414
	(0.109)	(0.105)		
CS performs valuable work for school (HT)	0.086	0.182*	0.359	4.286
	(0.100)	(0.105)		
CS is a good manager (HT)	0.029	0.115	0.463	4.257
	(0.105)	(0.100)		
Number of staff meetings by HT (T)	-0.059	0.656*	0.050	4.695
	(0.329)	(0.335)		
Number of staff meetings by CS (T)	0.028	0.280	0.335	2.060
	(0.230)	(0.283)		
CS provides useful feedback (T)	0.247***	0.289***	0.290	0.263
	(0.034)	(0.036)		
HT provides useful feedback HT (T)	0.167***	0.179***	0.728	0.519
	(0.038)	(0.037)		
HT feedback mentions something T did well (T)	0.181***	0.220***	0.488	0.536
	(0.055)	(0.054)		
HT feedback offers sugg. for improvement (T)	0.157***	0.219***	0.135	0.571
	(0.043)	(0.042)		
CS feedback offers sugg. for improvement (T)	0.286***	0.329***	0.454	0.252
	(0.052)	(0.053)		
CS feedback mentions something T did well (T)	0.283***	0.345***	0.291	0.255
	(0.052)	(0.052)		
T feel valued and appreciated (T)	0.019	0.045	0.493	0.430
	(0.040)	(0.039)		
HT has helped T become a better Teacher (T)	-0.017	0.036	0.211	0.286
	(0.040)	(0.041)		
CS has helped T become a better Teacher (T)	0.030	-0.000	0.408	0.221
	(0.035)	(0.036)		
HT is a good manager (T)	0.022	0.092*	0.187	0.307
	(0.053)	(0.053)		
Number of CS meetings with each HT in their circuit (CS)	0.000	-0.734	0.115	5.600
	(0.511)	(0.463)		
Number of CS meetings with all staff their circuit (CS)	-0.000	0.303	0.431	3.125
	(0.340)	(0.385)		
Number of CS visits to schools their circuit (CS)	-0.000	1.550***	0.000	4.679
	(0.299)	(0.369)		
Number of CS meetings with all HTs their circuit (CS)	-0.000	0.109	0.533	1.400
	(0.155)	(0.174)		
Suggestions to teachers frequency (CS)	0.000	-0.029	0.809	3.333
	(0.113)	(0.122)		
Strongly agree CS valuable mentor for teachers (CS)	0.000	0.095**	0.026	0.692
	(0.042)	(0.042)		
Strongly agree CS valuable mentor for headteachers (CS)	0.000	0.058	0.186	0.715
	(0.043)	(0.043)		
Strongly agree CS provides constructive feedback to teachers (CS)	0.000	0.097**	0.033	0.672
	(0.046)	(0.045)		
Strongly agree CS perform valuable work for school (CS)	-0.000	0.093	0.117	0.608
	(0.059)	(0.059)		
CS provides suggestions to HT (CS)	0.000	0.116	0.160	4.697
	(0.071)	(0.082)		
CS assessment of study school HT (strong agree) (CS)	0.014	-0.041*	0.042	0.029
	(0.028)	(0.023)		

Notes: Based on data collected at the first follow up. Columns 1 and 2: Coefficient estimates from Equation (1) for the listed component of the People

Table A5: Effects on Individual Components of Other Management Index

	Teacher Training + Mgmt Obs (1)	Teacher Training + Mgmt Obs + Ppl Skills (2)	P-value (3)	Control Mean (4)
An important part of HT job is to ensure teachers are held accountable (HT)	0.257***	0.206**	0.554	4.343
	(0.091)	(0.094)		
HT scheduled meetings with parents (HT)	-0.057	0.056	0.011	0.836
	(0.042)	(0.039)		
Teachers receive rewards for good performance (HT)	-0.086	0.039	0.109	0.700
	(0.078)	(0.079)		
School had student atendance records (HT)	0.029	-0.012	0.409	0.900
	(0.048)	(0.051)		
School had teacher atendance records (HT)	0.043	0.042	0.969	0.943
	(0.030)	(0.032)		
Schools uses teacher feedback to guide goals (HT)	0.086	0.149*	0.493	4.243
	(0.083)	(0.082)		
Teachers share new pratices with other teachers (HT)	0.357***	0.144	0.022	4.029
	(0.088)	(0.096)		
Teachers encourage students to approach them for supplemental help (HT)	0.214**	0.101	0.251	4.029
	(0.103)	(0.109)		
The school uses student scores to establish teachers' effectiveness (HT)	0.086	0.143	0.598	4.100
	(0.106)	(0.089)		
HT presence at spot checks (Spotcheck)	0.107	0.149**	0.525	0.421
	(0.066)	(0.068)		
An important part of CS job is to ensure teaching skills are improving (CS)	-0.000	0.038	0.593	4.742
	(0.071)	(0.071)		
An important part of CS job is to ensure teachers are held accountable (CS)	-0.000	0.035	0.644	4.667
	(0.072)	(0.075)		
Schools have goals/school development plan (CS)	0.000	0.223**	0.019	4.258
	(0.095)	(0.094)		
CS tried out new ideas in school (CS)	0.000	0.134	0.188	4.197
	(0.091)	(0.102)		
Schools use student scores to guide school goals (CS)	0.000	0.246**	0.018	4.152
	(0.105)	(0.103)		

Notes: Based on data collected at the first follow-up. Columns 1 and 2: Coefficient estimates from Equation (1) for the listed component of the Other Management index. All regressions include baseline management index and district fixed effects. Robust standard errors appear in parentheses. Column 1: Effect of Teacher Training + Management Observation. Column 2: Effect of the Teacher Training + Management Observation + People Skills intervention. Column 3: the p-value on the null hypothesis that the two effects are statistically equivalent. Column 4: the control group mean.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A6: Alternative Management Indices

Index: Reported by:	**Observational Practices Subordinate** (1)	**People Skills Subordinate** (2)	**Composite Management All Respondent Types** (3)
Teacher Training + Mgmt Obs	0.558***	0.285*	0.314**
	(0.135)	(0.157)	(0.148)
Teacher Training + Mgmt Obs + Ppl Skills	0.515***	0.623***	0.422***
	(0.125)	(0.149)	(0.142)
P-Value Same Effect	0.72	0.04	0.37
Observations	210	210	210
R^2	0.32	0.27	0.19
Mean Dep., Control	0.00	-0.00	0.00

Notes: The outcome variable in all regressions are standardized (by control group mean 0 and standard deviation 1) weighted indices of management outcomes. All regressions include baseline management index and district fixed effects. Standard errors are clustered at the school level. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A7: Effects on School Operations and Personnel Attendance at Unannounced Spot Checks

Attendance:	**School in Session** (1)	**Head Teacher Present** (2)	**Teacher Present** (3)
Teacher Training + Mgmt Obs	0.029*	0.111*	0.042
	(0.016)	(0.064)	(0.037)
Teacher Training + Mgmt Obs + Ppl Skills	0.028*	0.148**	0.028
	(0.016)	(0.064)	(0.042)
P-Value Same Effect	0.93	0.57	0.73
Observations	420	419	1,175
R^2	0.07	0.14	0.09
Mean Dep., Control	0.96	0.42	0.62

Notes: Regressions include district and survey round fixed effects. Standard errors clustered at the school level appear in parenthesis. Column 1: An indicator for whether or not students were having classes the day of the unannounced spot check. Columns 2 and 3: Additional covariates: respondent baseline attendance, age, gender, experience, age squared, and experience squared. Column 2: Indicators for whether a head teacher was present in school at the start of the unannounced spot checks. Column 3: Indicators for whether a teacher was present in school at the start of the unannounced spot checks. Additional covariate: Grade level taught.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A8: Classroom Activity Outcomes

Panel A: Classroom Activities Outcomes	**Any TLM Use** (1)	**Engaged Teacher** (2)	**Engaged Student** (3)	**Active Learning Index** (4)
Teacher Training + Mgmt Obs	0.192***	0.206***	0.180***	0.470***
	(0.032)	(0.025)	(0.029)	(0.060)
Teacher Training + Mgmt Obs + Ppl Skills	0.126***	0.172***	0.163***	0.362***
	(0.032)	(0.026)	(0.029)	(0.059)
P-Value Same Effect	0.03	0.21	0.52	0.06
Observations	2,462	2,462	2,462	2,462
R^2	0.08	0.10	0.08	0.10
Mean Dep., Control	0.48	0.46	0.48	0.00
Panel B: Classroom Activities Outcomes, Cond'l on Teacher Presence	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Obs	0.123***	0.148***	0.105***	0.370***
	(0.026)	(0.025)	(0.025)	(0.055)
Teacher Training + Mgmt Obs + Ppl Skills	0.054**	0.114***	0.107***	0.251***
	(0.026)	(0.025)	(0.023)	(0.051)
P-Value Same Effect	0.00	0.16	0.91	0.02
Observations	1,876	1,876	1,876	1,876
R^2	0.15	0.10	0.06	0.11
Mean Dep., Control	0.71	0.64	0.70	0.00

Notes: All regressions include controls for whether the observation is from the first or second spot check, the average percent of teachers present during the baseline survey, and strata (district) fixed effects. Standard errors in parentheses, clustered at the school level. Columns 1-3: Linear probability models. Column 1: TLM are teaching and learning materials. Column 4 repeats Table 2, column 2. The active learning index is based upon: teacher presence; any TLM use; teacher engaged; student express an idea. Panel A uses all observations. Panel B conditions on a teacher being present during the entire observation.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A9: DI Implementation Index Components

Dependent Variable:	**Did Leveling Exam**	**Had DI Handbook**	**Imp DI 4 Days in Last Week**	**Imp DI Today**	**Split in Levels**
	(1)	(2)	(3)	(4)	(5)
Teacher Training + Mgmt Obs	0.951***	0.740***	0.896***	0.779***	0.571***
	(0.015)	(0.038)	(0.023)	(0.031)	(0.047)
Teacher Training + Mgmt Obs + Ppl Skills	0.983***	0.767***	0.936***	0.797***	0.607***
	(0.010)	(0.036)	(0.019)	(0.031)	(0.047)
P Value Same Effect	0.06	0.59	0.17	0.67	0.54
Observations	210	210	210	210	210
R^2	0.97	0.72	0.92	0.78	0.55
Mean Dep., Control	0.00	0.00	0.00	0.00	0.00

Notes: Columns 1-4: School-level averages of teacher surveys. Column 1: Percent of teachers that reported doing the leveling exam. Column 2: Percent of teachers that had the DI Handbook. Column 3: Percent of teachers reported doing DI at least once in previous 5 school days. Column 4: Percent of teachers reporting that DI was done that day. Column 5: whether or not the enumerator observed that the class was split by levels, averaged to the school level. Table 2 Column 3 is at the spot-check level.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A10: Classroom Activities by Period

	Teacher in Classroom (1)	**Any TLM Use** (2)	**Engaged Teacher** (3)	**Engaged Student** (4)	**Active Learning Index** (5)
Panel A: First Period					
Teacher Training + Mgmt Obs	0.144***	0.200***	0.201***	0.180***	0.473***
	(0.035)	(0.036)	(0.037)	(0.038)	(0.078)
Teacher Training + Mgmt Obs + Ppl Skills	0.130***	0.152***	0.193***	0.173***	0.413***
	(0.040)	(0.040)	(0.038)	(0.040)	(0.082)
P-Value Same Effect	0.69	0.22	0.82	0.87	0.46
Observations	1,231	1,231	1,231	1,231	1,231
R^2	0.10	0.12	0.14	0.09	0.14
Mean Dep., Control	0.66	0.48	0.49	0.50	0.04
Panel B: Second Period					
Teacher Training + Mgmt Obs	0.117***	0.185***	0.210***	0.180***	0.466***
	(0.038)	(0.040)	(0.038)	(0.039)	(0.080)
Teacher Training + Mgmt Obs + Ppl Skills	0.093**	0.100**	0.152***	0.153***	0.311***
	(0.037)	(0.040)	(0.038)	(0.038)	(0.076)
P-Value Same Effect	0.48	0.02	0.14	0.44	0.05
Observations	1,231	1,231	1,231	1,231	1,231
R^2	0.08	0.08	0.10	0.08	0.09
Mean Dep., Control	0.70	0.47	0.43	0.45	-0.04

Notes: All regressions include controls for whether the observation is from the first or second spot check, the average percent of teachers present during the baseline survey, and strata (district) fixed effects. Panel A is the first period observation and Panel B is the second period observation. Standard errors in parentheses, clustered at the school level. See Table A6 for additional notes.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A11: Norms of Acceptable Behavior

	Teacher Opinions of Students	**Teacher Opinions of HTs**	**Teacher Opinions of CS**	**HT Opinions of Students**	**HT Opinions of HTs**	**HT Opinions of CS**
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Training + Mgmt Obs	-0.141	-0.254**	-0.203*	-0.201	-0.055	0.113
	(0.119)	(0.108)	(0.111)	(0.159)	(0.166)	(0.157)
Teacher Training + Mgmt Obs + Ppl Skills	-0.242*	-0.301***	-0.208	-0.197	-0.185	-0.037
	(0.132)	(0.115)	(0.131)	(0.146)	(0.152)	(0.157)
P-Value Same Effect	0.43	0.61	0.97	0.98	0.44	0.33
Observations	461	461	461	209	209	209
R^2	0.18	0.21	0.15	0.17	0.25	0.21
Mean Dep., Control	0.00	0.00	0.00	0.00	0.00	0.00

Notes: See Table 3. Column 1: Teachers' opinions of hypothetical students by teachers. Column 2: Teachers' opinions of hypothetical head teachers. Column 3: Teachers' opinions of hypothetical circuit supervisors. Column 4: Head teachers' opinions of hypothetical students. Column 5: Head teachers' opinions of hypothetical head teachers. Column 6: Head teachers' reports of hypothetical circuit supervisors.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A12: Work Stress and Burnout

	Teacher Stress & Burnout Index	**HT Stress & Burnout Index**
	(1)	(2)
Teacher Training + Mgmt Obs	0.042	0.043
	(0.081)	(0.118)
Teacher Training + Mgmt Obs + Ppl Skills	0.165**	-0.000
	(0.083)	(0.116)
P-Value Same Effect	0.18	0.70
Observations	1,849	621
R^2	0.09	0.19
Mean Dep., Control	0.01	-0.00

Notes: Based on data collected at the first follow-up of a index of 6 measures of agreement with various questions regarding work-related stress and burnout. All regressions include age, gender, experience, age squared, experience squared, and district and round fixed effects. Standard errors are clustered at the school level. Column 1: Teacher-reported. Column 2: Head teacher-reported. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A13: Additional Test Score Specifications

	Combined Score (1)	**Math Score** (2)	**English Score** (3)
Panel A: Limited Covariates			
Teacher Training + Mgmt Obs	0.106***	0.134***	0.067**
	(0.033)	(0.036)	(0.033)
Teacher Training + Mgmt Obs + Ppl Skills	0.081**	0.109***	0.050
	(0.037)	(0.039)	(0.037)
Panel B: Foundational Skills Only			
Teacher Training + Mgmt Obs	0.139***	0.183***	0.071***
	(0.027)	(0.034)	(0.022)
Teacher Training + Mgmt Obs + Ppl Skills	0.169***	0.203***	0.104***
	(0.029)	(0.035)	(0.027)
Panel C: Upper-Level Items Only			
Teacher Training + Mgmt Obs	0.075***	0.086***	0.045*
	(0.023)	(0.027)	(0.026)
Teacher Training + Mgmt Obs + Ppl Skills	0.054**	0.067**	0.033
	(0.026)	(0.030)	(0.027)
Panel D: Subset Asked of All Students			
Teacher Training + Mgmt Obs	0.118***	0.131***	0.075**
	(0.034)	(0.036)	(0.030)
Teacher Training + Mgmt Obs + Ppl Skills	0.114***	0.133***	0.064*
	(0.037)	(0.039)	(0.034)
Observations	5,608	5,608	5,608

Notes: Sample: all students available for the followup assessments. Outcomes: standardized latent ability of each student for combined, math, and English questions. All regressions include baseline controls for student age, age-squared, grade at baseline, baseline assessment scores, female, and strata (district) fixed effects. Standard errors in parentheses, clustered at the school level. Panel A: includes only baseline test score and strata as covariates. Panel B and C: Test questions are mutually exclusive and completely exhaustive. Panel B: Outcome is the score on the ASER-like items only. Panel C: Outcome is the score on upper level questions only. Panel D: Outcome is the score on questions asked of all students.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A14: Student Tested at Endline

	Tested (1)	**Tested** (2)
Teacher Training + Mgmt Obs	-0.003	-0.003
	(0.006)	(0.006)
Teacher Training + Mgmt Obs + Ppl Skills	-0.015**	-0.015**
	(0.007)	(0.007)
Baseline Test Score	0.006	0.004
	(0.003)	(0.005)
Teacher Training + Mgmt Obs X Baseline Test Score		0.006
		(0.007)
Teacher Training + Mgmt Obs + Ppl Skills X Baseline Test Score		-0.002
		(0.008)
P-Value Same Effect	0.06	0.06
Observations	5,893	5,893
R^2	0.02	0.02
Mean Dep., Control	0.96	0.96

Notes: The sample is all baseline students. All regressions include baseline test score, age, age-squared, and gender, district and round fixed effects. Standard errors in parentheses, clustered at the school level. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A15: Lee Bounds for Student Test Scores

	Combined Score		**Math Score**		**English Score**	
	Lower	**Upper**	**Lower**	**Upper**	**Lower**	**Upper**
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Training + Mgmt Obs	0.105***	0.106***	0.140***	0.138***	0.062***	0.064***
	(0.021)	(0.021)	(0.026)	(0.026)	(0.022)	(0.022)
Teacher Training + Mgmt Obs + Ppl Skills	0.101***	0.102***	0.124***	0.128***	0.072***	0.073***
	(0.024)	(0.024)	(0.030)	(0.029)	(0.024)	(0.024)
Observations	5,546	5,546	5,546	5,546	5,546	5,546
R^2	0.737	0.735	0.625	0.616	0.700	0.702

Notes: Lee (2009) bounding method applied to sample in Table 4. See additional notes in Table 4.

Table A16: Heterogeneity in Overall Student Test Score

	(1)	(2)	(3)	(4)	(5)
Teacher Training + Mgmt Obs	0.106***	0.098***	0.098***	0.110***	0.106***
	(0.021)	(0.027)	(0.030)	(0.028)	(0.031)
Teacher Training + Mgmt Obs + Ppl Skills	0.106***	0.091***	0.096***	0.123***	0.129***
	(0.024)	(0.029)	(0.028)	(0.033)	(0.035)
Teacher Training + Mgmt Obs X Baseline Test Score	0.011				
	(0.022)				
Teacher Training + Mgmt Obs + Ppl Skills X Baseline Test Score	0.002				
	(0.022)				
Teacher Training + Mgmt Obs X Female Student		0.023			
		(0.036)			
Teacher Training + Mgmt Obs + Ppl Skills X Female Student		0.034			
		(0.035)			
Teacher Training + Mgmt Obs X P4 Pupil at Baseline			0.020		
			(0.042)		
Teacher Training + Mgmt Obs + Ppl Skills X P4 Pupil at Baseline			0.023		
			(0.032)		
Teacher Training + Mgmt Obs X Either Parent Literate				-0.002	
				(0.034)	
Teacher Training + Mgmt Obs + Ppl Skills X Either Parent Literate				-0.029	
				(0.035)	
Teacher Training + Mgmt Obs X Above Median Amenities					0.002
					(0.043)
Teacher Training + Mgmt Obs + Ppl Skills X Above Median Amenities					-0.036
					(0.046)
Observations	5,608	5,608	5,608	5,608	5,608
R^2	0.746	0.743	0.743	0.744	0.746

Notes: See Table 4. The outcome variable in all regressions is the student's combined test score. Column 5: Amenity index calculated at the school level.

Table A17: Student Non-Cognitive Outcomes

	Student Absent From School	**School Reports Student Not Enrolled**	**Student Self-Reported Dropout**	**Fave Subject Math or Eng**	**Like Math or Eng Very Much**	**Want to Attend SHS**
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Training + Mgmt Obs	0.031**	0.028***	0.001	0.012	0.046***	-0.002
	(0.014)	(0.009)	(0.004)	(0.017)	(0.014)	(0.007)
Teacher Training + Mgmt Obs + Ppl Skills	-0.000	0.019**	0.004	-0.028	0.019	0.000
	(0.013)	(0.009)	(0.005)	(0.018)	(0.015)	(0.006)
P-Value Same Effect	0.02	0.38	0.40	0.02	0.05	0.72
Observations	11,569	5,893	5,608	5,608	5,608	5,550
R^2	0.03	0.02	0.02	0.06	0.05	0.03
Mean Dep., Control	0.17	0.06	0.01	0.79	0.79	0.97

Notes: All regressions include baseline test score, age, age-squared, and gender, district and round fixed effects. Standard errors clustered at the school level. Column 1: Whether the student was present the day of the unannounced spot check. The data could not be collected if classes were not in session. Data are one observation per student-wave. Column 2: Indicator equal to 1 if a teacher or head teacher reported that the student was no longer enrolled in the school. One observation per student. Column 3: Indicator equal to 1 if the student self-reported that they were no longer attending any school at the first follow-up survey. Col 4: Whether the child reports their favorite subject is math or English. Col 5: Whether the child reports they like Math or English very much. Col 6: Whether the child reports that they want to attend SHS. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A18: Management Outcomes in Year After Implementation: Individual Components of Observational Practices Index

	Teacher Training + Mgmt Obs (1)	**Teacher Training + Mgmt Obs + Ppl Skills** (2)	**p-value** (3)	**Mean for Control** ' (4)
Average Number of P4-P6 Classroom Obsevations (5 min) by HT (HT)	-0.324	-0.428	0.775	3.191
	(0.524)	(0.474)		
HT Classroom Observations (5 min) (T)	-0.089	0.207	0.476	3.133
	(0.446)	(0.427)		
CS Classroom Observations (5 min) (T)	0.143	0.355**	0.114	0.821
	(0.127)	(0.149)		
HT Classroom Observation Feedback (T)	0.155***	0.193***	0.414	0.626
	(0.053)	(0.051)		
CS Classroom Observation Feedback (T)	0.155***	0.202***	0.414	0.370
	(0.055)	(0.059)		
Average Number of P4-P6 Classroom Obsevations (30 min) by HT (HT)	0.601***	0.789***	0.380	0.870
	(0.208)	(0.216)		
HT Classroom Observations (30 min) (T)	0.578***	0.684***	0.581	0.971
	(0.214)	(0.168)		
CS Classroom Observations (30 min) (T)	0.201**	0.431***	0.026	0.353
	(0.084)	(0.097)		
Number of CS Classroom Obsevations (5 min) in study schools (CS)	0.011	0.541***	0.010	1.161
	(0.154)	(0.203)		
Number of school visits by CS (HT)	0.212	0.046	0.547	3.074
	(0.264)	(0.231)		
Number of CS visits to study school (CS)	0.215	0.236	0.952	3.233
	(0.301)	(0.324)		
Number of CS observations at study school (CS)	0.381*	0.734***	0.129	0.983
	(0.198)	(0.210)		

Notes: Based on data collected at the unannounced spot check in the year following the intervention. The outcome variables are standardized indices of management outcomes. All regressions include baseline management index and district fixed effects.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A19: Management Outcomes in Year After Implementation: Individual Components of People Skills Index

	Teacher Training + Mgmt Obs	Teacher Training + Mgmt Obs + Ppl Skills	p-value	Mean for Control
	(1)	(2)	(3)	(4)
HT provides constructive feedback (HT)	0.011	0.061	0.592	4.420
	(0.086)	(0.093)		
Number of staff meetings each term by HT (HT)	5.387**	9.799***	0.161	10.551
	(2.638)	(2.807)		
CS is a valuable mentor (HT)	0.058	0.143	0.463	4.391
	(0.106)	(0.107)		
CS works with HT to solves problems (HT)	0.054	0.051	0.984	4.435
	(0.111)	(0.117)		
CS performs valuable work for school (HT)	0.185	0.156	0.796	4.217
	(0.117)	(0.121)		
CS is a good manager (HT)	0.025	-0.028	0.674	4.290
	(0.117)	(0.106)		
CS provides useful feedback (T)	0.134**	0.184***	0.372	0.353
	(0.055)	(0.058)		
HT provides useful feedback HT (T)	0.172***	0.163***	0.856	0.585
	(0.052)	(0.051)		
HT feedback mentions something T did well (T)	0.182**	0.252***	0.352	0.504
	(0.077)	(0.081)		
HT feedback offers sugg. for improvement (T)	0.146*	0.236***	0.246	0.508
	(0.079)	(0.077)		
CS feedback offers sugg. for improvement (T)	0.142**	0.160**	0.827	0.277
	(0.066)	(0.077)		
CS feedback mentions something T did well (T)	0.168**	0.137*	0.718	0.284
	(0.071)	(0.077)		
T feel valued and appreciated (T)	0.041	0.035	0.907	0.367
	(0.050)	(0.050)		
HT has helped T become a better Teacher (T)	0.035	0.090*	0.300	0.295
	(0.050)	(0.052)		
CS has helped T become a better Teacher (T)	0.070	0.141***	0.139	0.200
	(0.047)	(0.047)		
HT is a good manager (T)	-0.042	-0.006	0.493	0.360
	(0.053)	(0.055)		
Number of CS meetings with all HTs their circuit (CS)	-0.013	0.947	0.404	1.489
	(0.424)	(1.158)		
Strongly agree CS provides constructive feedback to teachers (CS)	0.008	0.084	0.395	0.540
	(0.087)	(0.089)		
Strongly agree CS perform valuable work for school (CS)	0.000	0.080	0.345	0.603
	(0.083)	(0.085)		

Notes: Based on data collected at the unannounced spot check in the year following the intervention. The outcome variables are standardized indices of management outcomes. All regressions include baseline management index and district fixed effects.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A20: Heterogeneity in DI Implementation in Year After Implementation by Degree of Staff Retention

	DI Implementation			
	Low Teacher Retention Sample	High Teacher Retention Sample	Not Retained HT Sample	Retained HT Sample
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Obs	0.336	0.353***	0.445**	0.457***
	(0.204)	(0.090)	(0.199)	(0.087)
Teacher Training + Mgmt Obs + Ppl Skills	0.300	0.369***	0.428**	0.509***
	(0.182)	(0.126)	(0.188)	(0.108)
P-Value Same Effect	0.88	0.90	0.94	0.67
Observations	42	85	35	92
R^2	0.42	0.50	0.56	0.47
Mean Dep., Control	0.00	0.00	0.00	0.00

Notes: Based on data collected at the unannounced spot check in the year following the intervention. The outcome variable in each regression is the dummy variable indicating whether or not the students were divided by levels instead of grade during the unannounced spot check in the year after the intervention. Col 1: Sample is schools with below median teacher retention between the first and third spot checks. Col 2: Sample is schools with at or above median teacher retention between the first and third spot checks. Number is not equal to column 1 due to ties of the median value. Col 3: Sample is schools with a new head teacher between the first and third spot checks. Col 4: Sample is schools with the same head teacher between the first and third spot checks.

$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A21: Student Test Scores Two Years After Implementation: Heterogeneous Treatment Effects

	(1)	(2)	(3)	(4)	(5)
Teacher Training + Mgmt Obs	0.063**	0.086**	0.063*	0.033	0.051
	(0.026)	(0.034)	(0.035)	(0.038)	(0.042)
Teacher Training + Mgmt Obs + Ppl Skills	0.079***	0.068**	0.057*	0.074*	0.084**
	(0.024)	(0.032)	(0.031)	(0.038)	(0.036)
Teacher Training + Mgmt Obs X Baseline Test Score	0.036				
	(0.028)				
Teacher Training + Mgmt Obs +Ppl Mgmt X Baseline Test Score	-0.005				
	(0.024)				
Teacher Training + Mgmt Obs X Female Student		-0.051			
		(0.044)			
Teacher Training + Mgmt Obs +Ppl Skills X Female Student		0.019			
		(0.041)			
Teacher Training + Mgmt Obs X P4 Pupil at Baseline			-0.001		
			(0.043)		
Teacher Training + Mgmt Obs +Ppl Skills X P4 Pupil at Baseline			0.044		
			(0.037)		
Teacher Training + Mgmt Obs X Either Parent Literate				0.048	
				(0.047)	
Teacher Training + Mgmt Obs +Ppl Skills X Either Parent Literate				0.005	
				(0.046)	
Teacher Training + Mgmt Obs X Above Median Amenities					0.021
					(0.057)
Teacher Training + Mgmt Obs +Ppl Skills X Above Median Amenities					-0.005
					(0.048)
Observations	5,046	5,046	5,046	5,046	5,080
R^2	0.662	0.661	0.661	0.662	0.663

Notes: Sample is all students available for the assessments at the second follow-up. Regressions include controls for student age, age-squared, grade at baseline, baseline assessment scores, female, and strata (district) fixed effects. Standard errors clustered at the school level appear in parenthesis. Scores standardized relative to the baseline pooled sample. Column 5: Amenity index calculated at the school level.

Table A22: Student Non-Cognitive Outcomes Two Years After Implementation

	Currently Enrolled	**Want to Attend SHS**	**Fave Subject Is Math or English**	**Like Math or Eng Very Much**
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Obs	-0.017	0.003	0.054**	0.046***
	(0.013)	(0.011)	(0.023)	(0.015)
Teacher Training + Mgmt Obs + Ppl Skills	-0.004	0.006	0.013	0.026
	(0.013)	(0.012)	(0.023)	(0.016)
P-Value Same Effect	0.38	0.79	0.09	0.17
Observations	5,080	5,029	4,453	4,453
R^2	0.11	0.07	0.04	0.04
Mean Dep., Control	0.88	0.90	0.63	0.80

Notes: Sample is all students available for the assessments at the second follow-up. Regressions include controls for student age, age-squared, grade at baseline, baseline assessment scores, female, and strata (district) fixed effects. Standard errors clustered at the school level appear in parenthesis. Column 2 excludes students who responded "I don't know." Sample in Columns 3 and 4 excludes students who were no longer enrolled in school. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A23: Student Tested at Longer-Term Follow-Up

	Tested (1)	**Tested** (2)
Teacher Training + Mgmt Obs	-0.023*	-0.023*
	(0.012)	(0.012)
Teacher Training + Mgmt Obs + Ppl Skills	-0.032**	-0.032**
	(0.013)	(0.013)
Teacher Training + Mgmt Obs X Baseline Test Score		0.007
		(0.012)
Teacher Training + Mgmt Obs + Ppl Skills X Baseline Test Score		0.013
		(0.014)
Baseline Test Score	0.016***	0.010
	(0.006)	(0.009)
Observations	5,893	5,893
R^2	0.036	0.036
Mean of Dep., Control	0.880	0.880

Notes: The sample is all baseline students. All regressions include baseline test score, age, age-squared, and gender, district and round fixed effects. Standard errors in parentheses, clustered at the school level. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A24: Lee Bounds for Student Test Scores at Longer-Term Follow-Up

	Combined Score		**Math Score**		**English Score**	
	Lower	**Upper**	**Lower**	**Upper**	**Lower**	**Upper**
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Training + Mgmt Obs	0.055**	0.064**	0.073**	0.080**	0.047*	0.057**
	(0.026)	(0.026)	(0.032)	(0.031)	(0.027)	(0.027)
Teacher Training + Mgmt Obs + Ppl Skills	0.063**	0.075***	0.072**	0.082***	0.060**	0.073***
	(0.025)	(0.025)	(0.032)	(0.031)	(0.025)	(0.025)
Observations	4,993	4,993	4,993	4,993	4,993	4,993
R^2	0.654	0.653	0.548	0.542	0.625	0.628

Notes: Lee (2009) bounding method applied to sample in Table 7 at the Longer-Term Follow-Up. See additional notes in Table 7.