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

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-0003536. 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

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

ABSTRACT

To improve public services, public sector managers must encourage reticent civil servants to enact effective reforms. We show through a randomized controlled trial that school principals, i.e., school mangers, can act as leaders to improve Instructional Management (0.3SD) and student learning (0.11SD) with existing systems and personnel. Additional management training improved People 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 commitment, acting as leaders.

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

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

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

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

1 Introduction

Public service productivity and service provision in developing countries are often deficient. In particular, the government schooling sector's lack of productivity is acute as millions of students are in school but not learning, i.e., the learning crisis (World Bank, 2018). Randomized controlled trials (RCTs) have improved learning (e.g., Banerjee et al., 2007; Banerjee et al., 2017; Duflo et al., 2022). Yet, civil servants have been reticent to implement these programs. Unlike the private sector, civil servants cannot be compelled by the private sector management tools of pay, promotion, and termination to increase their effort levels (Bloom et al., 2013; Bloom et al. 2018; Bruhn et al. 2018). Nevertheless, the public expects, and deserves, high quality public service. Since public sector managers do not have these tools of compulsion, they must act as leaders, eliciting others to follow. We show through an RCT that public sector managers can act as leaders to improve student learning with existing systems and personnel.

We partnered with the Ghanaian government to apply insights from the organizational, management, and education literatures to improve productivity 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 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., 2022).[1] Yet, getting existing teachers in existing systems to implement the program has been elusive in Ghana and India because of the educator sector's weak managerial and oversight structures, leading to implementation rates of around 5% (Banerjee et al. 2017; Duflo et al., 2022). Are public sector managers the key to reform?

We show through a conceptual framework that public sector workers increase their effort when managers act as leaders through the use of credible, costly effort signals that convince

[1]Differentiated Instruction is sometimes known as Teaching at the Right Level (TaRL), Targeted Instruction (TI), or Differentiated Learning (DL).

workers of the return to and importance of their own effort, in the "lead by example" style of Hermalin (1998). Public sector managers might further increase school productivity through support and mentorship only if it leads to additional worker effort.

Our 210 school, three-armed RCT in Ghana demonstrates the power of managers acting as leaders to increase teacher effort and student learning. The Teacher Training + Management Effort Treatment (Treatment 1) trained teachers in Differentiated Instruction (DI) and gave managers (i.e., school principals and circuit supervisors) costly effort tasks (primarily conducting a classroom observation and related form). These tasks credibly signaled to teachers that the new DI classroom activities were important, valued, and would increase student success, like a leader uses an effort signal to elicit change from followers. The Teacher Training + Management Effort + People Management Treatment (Treatment 2) included all the elements of Treatment 1 plus an additional training for school managers on People Management – improving interpersonal relationships through mentoring and support.[2] 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. Our conceptual framework predicts that the costly and visible management effort in Treatments 1 and 2 will induce teachers to exert effort on DI, improving student learning. The People Management training in Treatment 2 will only further increase student learning if it leads to additional teacher effort.

The two interventions facilitated the costly effort signal and improved school management, classroom teaching quality, and student learning. The gains persisted after the end of the intervention. Each treatment increased by 30 percentage points the likelihood that a teacher was observed for at least 30 minutes – a costly effort signal was sent. Both interventions' emphasis on implementing DI and higher quality and more frequent classroom observations increased Instructional Management by 0.3 standard deviations (SD) relative to the control group. The additional mentoring and support training in the People Man-

[2]In the private sector, People Management 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.

agement arm improved People Management quality relative to both control and Teaching Training + Management Effort alone schools.[3] Both treatments increased teacher engagement with their classrooms and led to equal, high Differentiated Instruction implementation. 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, and were still implementing DI without continued training or materials.

The differential improvements in People Management but similar changes in Instructional Management and classroom practices led to equal changes to student learning across the two interventions – about 0.11 SD, an additional one third of a year of schooling. Students from treatment schools continued to have test scores that were higher than their control school peers two years after the end of the intervention, through Covid-related school closures and students graduating from primary to junior high school.

The importance of Instructional Management and not People Management is further supported by a mediation analysis: increases in student test scores were positively related to increases in Instructional Management and unrelated to increases in People Management. 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.

In addition to its policy relevance, as national governments in Africa and state governments in India 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 importance of combining the literatures on management and leadership, treating public sector managers as leaders capable of inspiring change through a costly effort signal.

[3]We adapt two measures of management quality from the internationally validated Development World Management Survey (D-WMS) – Instructional Management, i.e., instructional policies and how instructional practices are managed, and People Management, i.e., the quality of interactions between teachers and their managers (Bloom et al., 2015; Lemos and Scur, 2016).Management of schools includes more than Instructional and People Management. See more details on what is included in each type of management in Section 6.3.

Theoretical models tout the importance of leadership among workers (Hermalin 1998), in social movements (Loeper et al., 2014), to create social norms (Acemoglu and Jackson 2105), and in investments (Akerlof and Holden 2016). Empirically, the importance of leadership has been shown in workers' unions (Boudreau et al., 2022). We apply these lessons to public sector managers who must inspire but cannot compel workers (front line civil servants) to undertake change, as the organization of public sectors leaves little scope for any layers of management to compel worker change. Previous studies that included training existing teachers and materials for Differentiated Instruction, but did not have a leader, led to no increase in test scores after one year (Banerjee et al., 2017; Duflo et al., 2022). With the inclusion of a costly effort signal that caused managers to act as leaders, our effect sizes are similar in magnitude to the version in which NGO personnel were leaders (Banerjee et al., 2017). Further, the management and implementation effects persisted one year after the program. 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.[4]

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 found 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).[5] The breakdown in this relationship is in contrast to the correlations between management quality and productivity in this study (a 1SD difference in management quality is correlated with a 0.15SD difference in student test scores), previous correlations from the private (e.g., Bloom and Van Reenen 2007; Hsieh and Klenow 2009; see summary in Bloom and Van Reenen 2011; Bloom et al., 2019; Gibbons and Henderson 2012) and public sectors (Rasul and Rogger 2018;

[4]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).

[5]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.

Tsai et al. 2015; Bloom et al., 2015; Lemos and Scur 2016; Crawfurd 2017; Lemos et al. 2021), and evidence on management improvements increasing productivity in the private sector (Bloom et al. 2013; Bloom et al. 2018; Bruhn et al. 2018). Improving management in the public sector in developing countries is a nascent field with no papers appearing prior to 2017 (Finan et al., 2017).[6] By separating management into multiple components, we show that not all management is equally important in the production of learning – better Instructional Management changed classroom practices and increased learning while additional improvements in People Management did neither. We also show that existing public sector school management can be improved and can increase student learning.

Third, the entire intervention from material design to implementation was government led – existing systems can scale effective programs with sufficient political will. 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 and across schooling levels. Starting in 2022, the Teacher Training + Management Effort intervention was being scaled to 10,000 schools in Ghana.

[6]The evidence in developed country schools is scant and often involves massive (300 hour) time investments (Fryer 2017). Recent approaches in developing countries outside the existing systems include outsourcing management to private sector firms (Romero, Sandefur, and Sandholtz 2020) or creating additional layers parallel to the existing system (Cilliers et al. 2020a; Cilliers et al. 2020b).

2 Background

2.1 The Ghanaian Education Sector

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., 2022). 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.[7]

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 terms of teacher employment.[8] Each school belongs to a circuit of approximately 8 geographically proximate schools, overseen by a circuit supervisor.[9] 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

[7]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. 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 Covid19 beyond their school duties.

[8]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 principal-specific training prior to this intervention.

[9]As with head teachers, circuit supervisors are almost always previous teachers and, like head teachers, receive almost no additional formal training.

teaching, usually a 5 minute peek through a door or window.

2.2 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. It improves test scores when educators implement it, but even when trained, teachers have been reticent to implement it in India and Ghana (Banerjee et al. 2007; Banerjee et al. 2017; Duflo et al., 2022). The degree of implementation is a key determinant of the heterogeneity of effect sizes across studies and contexts (Angrist and Meager 2023). As of 2023, it was part of education policy in sixteen countries across Africa and South Asia.

3 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 al. 2017, Duflo et al., 2022). With existing systems and personnel, what can increase the likelihood that civil servants exert effort to implement programs that increase productivity?

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) can induce greater worker (teacher) effort by acting as leaders. As with a leader, head teachers are similar to workers (almost all are 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, this means 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,

$$A = \theta\, f(e_1, ..., e_n, e_{HT}) + \varepsilon \tag{1}$$

where $-1 \leq \theta \leq 1$ and θ is not known to teachers.

Teachers have beliefs about θ based on their previous experiences and head teachers' behaviors. DI is new to these teachers–it was not how they were taught nor how they were taught to teach. Let $\tilde{\theta}$ reflect teachers' beliefs about θ, $-1 \leq \tilde{\theta} \leq 1$. Their skepticism about new teaching methods will cause their beliefs to understate the true return to DI, therefore $\tilde{\theta} \leq \theta$.

Teachers get utility from their fixed wages, student achievement, 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

$$\max_{e} U\,(e_n) = w + A(\tilde{\theta}, e_1, ..., e_n, e_{HT}) - d\,(e_n) + F\,(V_s) \tag{2}$$

where w is the teacher's fixed wage, $A(.)$ is the achievement function from above but with a teacher's belief about the productivity 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.[10] The disutility function can be seen as a teacher's forgone

[10] As with many civil servants, teachers' wages, w, do not depend on actual 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 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 teachers of subsequent grade levels for poor student performance. 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 performance.

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 student 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 $A\left(.\right)$ as specified above, a scaling parameter γ and other school amenities, $F\left(V_s\right)$.[11] He gets disutility from exerting DI effort, $d\left(e_{HT}\right)$.

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. 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 example," they can induce teachers to update their belief about θ.

Following Hermalin (1998), HTs choose their effort level first. Teachers observe this effort level, which can inform 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.[12] If this HT effort

[11] As with teachers, head teachers' wages do not depend on student 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 performance affects head teacher utility without affecting their wages, for example because of intrinsic motivation, students being easier to manage when they are learning, it could affect future school assignments, or they could be chided by parents or school officials for poor student performance. Both A and e may eventually affect w if they lead to a head teacher promotion to district leadership.

[12] 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 resulting 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.

is not readily ascertained by the teachers, it conveys no information, and teachers do not update $\tilde{\theta}$.

Not all management interventions provide tools to allow managers to act as leaders. The leadership framework predicts that management interventions that visibly increase e_{HT} will increase teacher effort and thus student achievement. Management interventions that do not visibly increase e_{HT} will not affect teacher effort and thus lead to no change in student test scores. This second type of management intervention might increase V_s, and thus teacher utility, but will not affect student achievement.

Our interventions included visible elements of costly managerial effort – attending the training with the teachers and an observation form that simultaneously provided a simple way for HT to demonstrate their increased effort, i.e., a costly effort signal, and reinforced the most important aspects of DI implementation. To complete the observation form, head teachers spent at least 30 minutes in a teachers' classroom, noting what was done well and areas for improvement.[13] 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 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.

Additional managerial practices, like those conveyed in the People Management training, will only increase student learning if they affect the amount of teacher effort. Any intervention that improves V_s increases teacher utility but does not affect teacher effort.

4 Intervention

The interventions applied the theory of leadership to induce teachers to engage in Differentiated Instruction, a primary school pedagogy reform. Both interventions were designed

[13] One of the head teachers' jobs had always been teacher observation, but based on our data this typically involved a 5 minute peek through the door. The observation form was a more involved undertaking.

as management level interventions that could be implemented 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 section describes the two interventions. The initial trainings occurred prior to the start of the 2018-2019 academic year.

4.1 Treatment 1 – Teacher Training + Management Effort

Treatment 1 trained teachers and provided tools and opportunities for managers to use their own effort to signal the positive return to implementing the program in schools.

Teacher Training

Teachers of grades 4 through 6 received in-service training on how to implement Differentiated Instruction and teaching and learning materials to support dividing students by learning level for part of the school day and creating a more student focused, active learning practice. Students in grades 4 through 6 had DI time with their learning level instead of grade level one hour each day, four days per week, eight weeks per term, i.e., partial day tracking. One teacher was assigned each learning level, e.g., the grade 4 teacher might teach learning level 1.[14] Because learners from all three grades grade 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. Appendix Section A.1 contains additional intervention details.

Teachers decided how much effort, e_n to exert on this new, to them, teaching method based on $\tilde{\theta}$, their estimate of θ. Head teacher and circuit supervisor behavior could affect the accuracy of this estimate.

Management Effort

As the optimal teacher effort is increasing in θ, credible signals, i.e., those that were not pronouncements but required effort, about θ from head teachers and circuit supervisors would increase teacher effort on DI. This intervention embedded multiple opportunities for

[14]Students could be in a different math and English levels and could move between levels during the school year as their progression warranted. Teachers tended to stay with the same learning level for the entire year.

head teachers and circuit supervisors to "lead by example" and increase teachers' estimate of θ. The most costly and likely salient way that teachers observed this effort was the head teachers' and circuit supervisors' 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. They 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 would increase student learning. This replaced an existing evaluation form that focused on whether students completed exercise books and were well behaved. Head teachers and circuit supervisors were additionally invited to the teacher training to learn how teachers should implement DI and support teachers in that implementation. Finally, a national level monitoring team displayed their own costly effort by visiting each study district 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.[15]

4.2 Treatment 2 – Teacher Training + Management Effort + People Management

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

People Management

Positive relationships between workers and managers are another component of high quality management. One of the few other management techniques available to public sector mangers are their interpersonal relationships with workers. 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

[15]National level monitors visit schools as regular policy, but less frequently, averaging about 16 percent of schools per year. During the intervention, they visited 88 percent of treatment schools.

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 for support.[16] 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 head teacher and teacher relationship and differentiated instruction components and dates.[17] Whether improved manager-worker relations are a costly managerial effort signal that induces additional teacher effort or an amenity that is valued by teachers, V_s, but does not affect their effort choice is an empirical question.

5 Empirical Strategy

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

$$y_{is} = \alpha + \beta_1 T1_s + \beta_2 T2_s + \Gamma' X_{is} + \varepsilon_{is} \tag{3}$$

[16] 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.

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

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 Effort treatment, $T2_s$ is an indicator equal to 1 for schools in the Teacher Training + Management Effort + People Management 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.[18]

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 Effort relative to the status quo through the estimate of β_1. The estimate of β_2 captures the total effect of Teacher Training + Management Effort + People Management relative to the status quo. The difference between the estimates of $\beta_1$and β_2 provides the differential effect of People Management.[19] We provide further insight into the relative and absolute importance of the two interventions through a mediation analysis.

6 Sample Selection and Data

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

[18]In student-level regressions, to improve precision we additionally control for student age and age-squared, baseline grade level, and an indicator variable for being female. 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.

[19]We do not include an intervention that is teacher training alone as both Duflo et al., (2022) 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.

6.1 Sample Selection and Randomization Procedure

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

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 Effort intervention and 70 circuit supervisors to receive the Management Effort + People Management intervention.[21] [22] We randomly selected two eligible schools from each circuit.[23] Within each of the Management Effort circuits, we randomly assigned one selected school to be a control school and the other to be a Teacher Training + Management Effort school. To be consistent across treatments, within each of the Management Effort + People Management circuits, we randomly assigned one selected school to receive the Teacher Training + Management Effort + People Management 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 Train-

[20]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.

[21]These districts contained 145 circuits. Ex ante we excluded two circuits from the study for piloting and three circuits because of too few eligible schools, leaving 140 study circuits. See below for details on eligibility.

[22]Circuit supervisors of control schools received the Management Effort 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., 2022). 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 UNICEF districts, the available budget, and existing estimates.

[23]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 characteristics.

ing + Management Effort, Teacher Training + Management Effort + People Management, and control.[24] 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 Effort compared to business-as-usual, and the impact of Teacher Training + Management Effort + People Management compared to business-as-usual. It also allows us to compare the differential effect of adding the People Management intervention relative to the Teacher Training + Management Effort intervention alone. In all cases we establish both effectiveness and cost-effectiveness.

6.2 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 Covid19 school closures) spot check round the year after the intervention, and a follow-up two years after the end of implementation. 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.[25] We surveyed all 140 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

[24]This study is similar to a fully cross-randomized design, although we do not have any schools that only received the People Management training without Teacher Training + Management Effort 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 of Teacher Training alone (Banerjee et al., 2017 and Duflo et al., 2022) and Management training alone (Muralidharan and Singh 2020; Ganimian and Freel 2021; Romero et al., 2022).

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

the management measures appear in Section 6.3. We further surveyed and tested a random sample of 15 students from P4 and 15 students from P5.[26] 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 program take-up and implementation and management. 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 schools were supposed to be visited in each spot check, but the third spot check was disrupted by the Covid pandemic.[27]

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

[26]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.

[27]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.

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.

6.3 Management Indices

We used the previously validated Development World Management Survey (D-WMS) instrument as a framework for measuring school managerial practices (Lemos and Scur 2016). Because of the interventions' foci on affecting classroom instruction and interactions between managers and teachers, we focus on Instructional Management, People Management, and Other Management–creating an index to capture each, following Anderson (2008). Teachers', head teachers', and circuit supervisors' responses and direct observation were used to create the indices. The positive correlations between each index and student test scores in the control group appear in Appendix Figure A1.

Instructional Management

The Instructional Management Index captures the elements of the D-WMS related to the management of instruction and classroom practices, elements that could have been affected by Teacher Training + Management Effort, such as the number of classroom observations by supervisors, standardization of instructional planning processes, personalization of instruction and learning, data-driven planning and student transition, and performance tracking. All components of the index appear in Appendix Table A1.

People Management

The People Management Index captures the elements of the D-WMS related to how supervisors and subordinates interact measuring performance review and performance dialogue, the relationship between subordinates and supervisors, and the degree to which teachers received mentoring, coaching, and feedback to improve their teaching practices – elements affected by the People Management training. The full list of components appears

in Appendix Table A2.

Other Management

The final index includes other aspects of management, which might improve student learning but are not directly related to Instructional or People Management including questions on 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. All aspects of this index appear in Appendix Table A3.

6.4 Summary Statistics and Baseline Balance

Appendix Table A1 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.

A few statistics, which are statistically equal 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.

7 Results

We start with estimating the effects of the intervention on management. 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.

7.1 Effort Signal and Management Practices

Both interventions caused managers to spend more time observing their teachers, i.e., providing a costly effort signal, which should have increased teachers estimates of the return to DI, θ. Teachers in treatment schools were about 35 percentage points more likely to have been observed for at least 30 minutes by either their circuit supervisor or head teacher, about doubling the control group mean of 39 percent (Table 1, column 1).

The interventions further improved the management of classrooms and instructional practices, increasing the Instructional Management score by about 0.3SD relative to a control group management score standardized to mean 0 and standard deviation of 1 (column 2).[28] The effects of the interventions on the individual components of the Instructional Management Index appear in Appendix Table A1.

[Table 1 about here]

The effect of the interventions on People Management is only statistically significant for the intervention that included specific, additional training on People Management, increasing that measure by 0.7SD (column 3). We reject the equality of the coefficients across the two interventions at the 0.02 level. The effects of the interventions on the individual components of the People Management Index appear in Appendix Table A2.

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

[28]The p-value for the Teacher Training + Management Effort is 0.058.

Effort and 0.6SD for the Teacher Training + Management Effort + People Management – with the effect of the intervention that included People Management statistically larger at the 10 percent level (p-value=0.07; column 4). Appendix Table A3 contains the estimates of the effect on each component of the Other Management index separately.

Therefore, managers in both treatments sent a costly, credible effort signal by spending the necessary time in classrooms to complete the DI observation tool and increased schools' Instructional Management quality. These changes were statistically equivalent across the two interventions. The schools that received the additional People Management training had higher People Management quality than both the control and the Teacher Training + Management Effort schools. Both interventions further had spillovers into other aspects of management, more so for the intervention that included People Management.[29]

7.2 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, we conducted two unannounced spot checks when schools were supposed to be holding normal school operations and DI lessons. Relative to a control group mean of 96%, the interventions increased the likelihood that school was in session by 3 percentage points (Table 2, column 1). To lead by example, head teachers must be present. Head teachers were 11 (Teacher Training + Management Effort) or 15 (Teacher Training + Management Effort + People Management) percentage points (26 and 37 percent) more likely to be present relative to a control group mean of 42 percent (column 2). On average only 62 percent of teachers across all three arms were present at the start of our spot-check visits. We do not find any effect on

[29]We confirm in Appendix Table A5 that these are not enumerator demand effects – the effects are similar, including the differential effects for the People Management arm, when the indices remove the self-reported outcomes and rely only on subordinate reports. Teachers (the subordinates) received the same training in both interventions.

teacher attendance (column 3).[30] The two interventions were equally effective at improving school operations.

[Table 2 about here]

Classroom Activities

Classroom instruction improved as teachers aligned their practices with those from the teacher training (Table 3), demonstrating that they internalized the signals sent by their managers on the return to DI effort. Teachers were more likely to be in the classroom (11 percentage points, column 1) and scored more highly on an index of active learning techniques, techniques taught in the Teacher Training (increase in 0.47SD for Teacher Training + Management Effort, 0.36SD for Teacher Training + Management Effort+People Management, statistically different, column 2).[31] Therefore, classrooms in Teacher Training + Management Effort schools became more active than those in Teacher Training + Management Effort + People Management schools and both had more active learning than in control schools. [32] Active Learning improved more in the intervention without People Management training perhaps because head teachers only focused on DI and not other aspects of management highlighted in People Management training.

[Table 3 about here]

In addition to using active pedagogy, DI training instructed teachers to divide students by learning levels for one period each day. We observed about 60 percent of schools teaching

[30] 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.

[31] The effects on the individual index components appear in Appendix Table A6 Panel A. The results are similar if we condition on a teacher being present in the classroom (Table A6, Panel B).

[32] As these classroom observations occurred during the first two instructional periods of the school day and DI was only one period long, some of the improved pedagogy occurred outside of DI time. The training recommended implementing differentiated instruction in either of the first two periods of the day, on Tuesdays through Fridays, and starting the third week of the term. We timed our visits to match these guidelines. 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. In Appendix Table A7 we estimate the outcomes in Table 3 separately by Period. 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.

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). This likely understates the true adherence as 22 percent of head teachers reported that their school conducted DI lessons at times other than the first two periods of the day. Column 4 combines this outcome with four additional self-reported measures of DI implementation – 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 – into a single index finding that the average school across both treatment arms was implementing 80 percent of the program.[33] Teachers exerted effort in implementing DI.

Norms and Stress

The interventions further changed norms around the acceptability of shirking–reinforcing that teachers updated their beliefs about the return to teacher effort. 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. Table 4 contains these results, replacing the dependent variable in Equation 1 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 Effort) or 0.4SD (Teacher Training + Management Effort + People Management) (column 2). Head teachers similarly viewed shirking by teachers more harshly – about 0.3SD (column 3). While the point values are negative on the overall index of head teachers' views, it is not statistically significant.[34]

[33]Appendix Table A8 shows the effect of the interventions each component of the index—in all cases the point values are larger for the Teacher Training + Management Effort + People Management intervention but only statistically different (p-value=0.06) for completing the leveling.

[34]Appendix Table A9 presents the opinions about the other agents at the school separately.

[Table 4 about here]

One concern could be that the additional effort by teachers found in Table 3 and head teachers in Tables 1 and 2 would not be sustained due to increased stress and burnout. Head teachers in treatment schools exhibit similar levels of stress and burnout to their control school peers as do teachers in the Teacher Training + Management Effort intervention (Appendix Table A10, columns 1 and 2). Teachers in the Teacher Training + Management Effort + People Management intervention reported higher levels of stress and burnout perhaps because head teachers' implementations of improved interpersonal relationship training felt like additional monitoring instead of support (Appendix Table A10, column 1).

7.3 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 5 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).[35] Over this same period, about one school year, control group students learned about 0.3SD. Therefore, these test score increases are 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 Management differentially changed in the Teacher Training + Management Effort + People Management intervention, the test score improvements were almost identical across

[35] Appendix Table A11 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. 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).

the two interventions.[36]

[Table 5 about here]

We tested for heterogeneity in program impacts on students' test scores by students' baseline test score, gender, grade level at baseline, parents' literacy, and a school amenity index, finding no heterogeneous effects (Appendix Table A12).[37] Appendix Figure A2 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 (Banerjee et al. 2007; Banerjee et al. 2017; Duflo et al., 2022).

Non-cognitive Outcomes

Dividing students by learning levels might discourage students and affect a student's 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 A15).[38] The interventions did not decrease enthusiasm about school or future aspirations (Appendix Table A16). Instead, the Teacher Training + Management Effort

[36]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 Effort + People Management arm than the control group but this is not differential by both test score and treatment (Appendix Table A13). Nevertheless, in Appendix Table A14 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.

[37]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.

[38]Students in the Teacher Training + Management Effort 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 A15, 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 school since they were to test and record scores for students each term. When reached directly, students across all treatment arms are equally likely to report that they were still attending school (column 3).

intervention increased the likelihood that students reported liking English or Math, i.e., the two intervention subjects, very much by 5 percentage points, a higher rate than the other intervention or the control group. Overall, we find no evidence of student discouragement.

7.4 Persistent Effects

Consistent with head teachers continuing to act as leaders and teachers permanently updating their beliefs about the return to their Differentiated Instruction effort, the interventions improved management and teachers continued to implement Differentiated Instruction even after the treatments concluded.[39] Treatment students' test scores were higher than control students 2 years after the end of the intervention.

Management

Management improvements persisted into the year after the intervention. We find a positive point value effect on Instructional Management, statistically different from 0 for the Teacher Training + Management Effort + People Management schools (Table 6, column 1). People Management quality was still statistically significantly higher in both types of treatment schools than control schools (column 2).[40] Across both management indices, the intervention that included People Management 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 teacher satisfaction with their current postings. The Teacher Training + Management Effort 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. Teacher Training + Management Effort + People Management reduced head teacher turnover by 11 percentage points and teacher turnover by 9 percentage points (columns 3

[39]In the year after the intervention, GES instructed schools to continue with Differentiated Instruction into the next academic year (AY2019-2020) and invited one teacher per school to a brief refresher training but did not distribute any materials to use in the new school year.

[40]Effects on the individual components of the indices appear in Appendix Tables A17 and A18.

and 4). Head teacher turnover could have tempered the persistent effects on management practices.

For both retention measures we reject that the coefficients across the interventions were the same – the additional People Management led to more head teachers remaining than either the control or non-People Management intervention. In contrast, the highest proportion of teachers were retained in the Teacher Training + Management Effort schools. 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 teacher did well, yet teachers in the People Management Intervention also reported higher levels of job related stress and burnout (Appendix Tables A3 and A11). These countervailing forces likely led to the increased teacher retention in the People Management intervention relative to the control group, but lower than in the non-People Management Intervention.

[Table 6 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 effectiveness of Differentiated Instruction, $\tilde{\theta}$ in the conceptual framework. About 40 percent of schools were still dividing their students by learning levels for one of the first two periods of the day (Table 7, column 1). While lower than during the intervention year, this level of implementation is about 6 times higher than was observed previously in Ghana during the intervention years when management was not engaged (Duflo et al., 2022). 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.

[Table 7 about here]

We test for heterogeneity by the degree of teacher turnover and whether the head teacher was still at the school. Schools across both treatments were still implementing DI during 35 percent or more of observations regardless of teacher turnover or head teacher departure (Appendix Table A19).

Student Achievement

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 8). Despite only a year long intervention, an enormous schooling disruption from Covid-19, and changing schools and schooling levels (students should have been in JHS1 and JHS2), the interventions increased student learning two years after the intervention by 0.06SD for students who had been in the Teacher Training + Management Effort schools and 0.08SD for students who had been in the Teacher Training + Management Effort + People Management 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.[41] 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. 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 A20). Students from the Teacher Training + Management Effort schools were more likely than both the other intervention and the control group to say that their favorite subject was Math or English and more likely than the control group to report liking math or English very much (Appendix Table A21). We do not find differential selection into test taking by the interaction of treatment status and baseline test score, but provide Lee (2008) bounds nevertheless (Appendix Tables A22 and A23).

[41] 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.

8 Cost Effectiveness

We use the ingredients method to assess cost effectiveness (Hirji et al., 2023). Teacher Training + Management Effort cost $41 per student and the Teacher Training + Management Effort + People Management cost $74 per student. As the achievement effect sizes were almost exactly equal across the two interventions, the Teacher Training + Management Effort intervention was clearly more cost effective. To scale this relative to the effect size per $100, this would be 0.26SD per $100 for Teacher Training + Management Effort and 0.14SD per $100 for Teacher Training + Management Effort + People Management. To scale this as a price per year of learning in this context, Teacher Training + Management Effort costs $125 per student per year of learning and Teacher Training + Management Effort + People Management costs $228 per student per year of learning.[42] The achievement effects persisted for at least two years with minimal additional costs, increasing the overall cost effectiveness.

9 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. 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 Management 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

[42]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.

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 Management differentially increased two of the three measures of management, but did not increase test scores beyond the Teacher Training + Management Effort 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 Instructional Management being more important for student learning than People 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 Management 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 Management relative to the treatment schools without it.

To further explore the relationship between Instructional Management, People Management, 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 Instructional Management, People Management, 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 3 presents the results of this exercise, demonstrating that there is a larger correlation between Instructional Management and test score improvements

than People Management and test score improvements.[43] 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 Management had the same test score increases as those with smaller gains in People Management.

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. Teachers continued to implement the program, both Instructional and People Management 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 Effort 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.

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

[43]The OLS regression coefficient linking test scores and instructional management is 0.0140 (p=0.000) while the OLS regression coefficient linking test scores and people management is 0.0003 (p=0.844).

a second intervention that also included People Management 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 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.

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," Working Paper 24249, National Bureau of Economic Research January 2018.

_ , 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.

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.

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.

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 _ , "Experimental Evidence on Alternative Policies to Increase Learning at Scale," Working Paper 27298, National Bureau of Economic Research May 2022.

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.

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.

Gibbons, Robert 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.

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.

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.

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

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.

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.

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), which targeted students in grades 1 through 3. They then modified these existing resources and designed and developed new materials as appropriate for older learners. 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 in hypothetical settings.

The main teacher training occurred for 5 days prior to the start of the 2018-2019 academic year with shorter refresher trainings prior to the start of each term, 3 day 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 Management Material Development

The National Inspectorate Board (NIB), the board that oversees school inspection and evaluation, developed the circuit supervisors' and head teachers' People Management 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 Management Training

The People Management 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 suggested that schools implement DI during the first period of the day. Schools could implement it at a time that worked best for them–43 percent selected Period 1, 33 percent Period 2, and 22 percent 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 (2022), 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 (2022), 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.

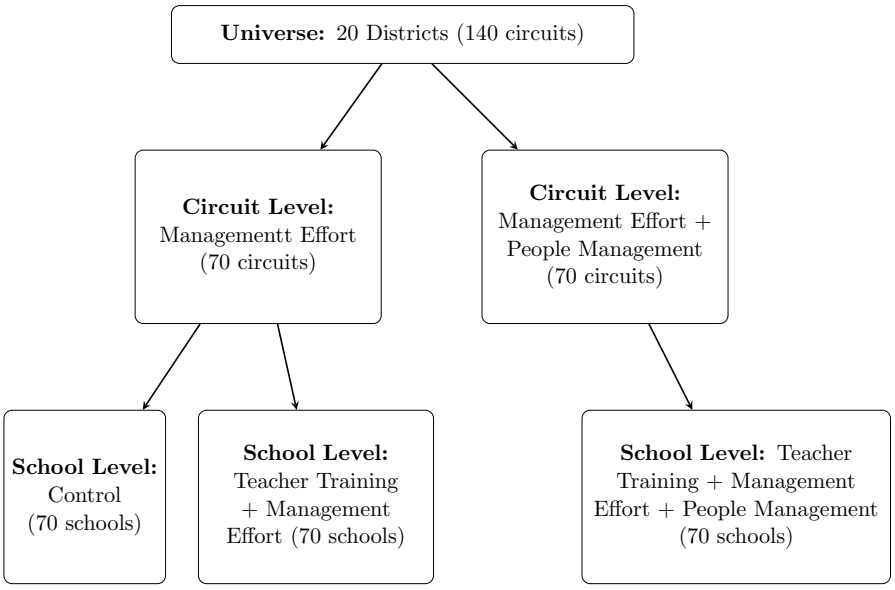
[Appendix Figure A2 about here]

A.3 Additional Tables

Tables A1 through A23 contain additional supporting materials.

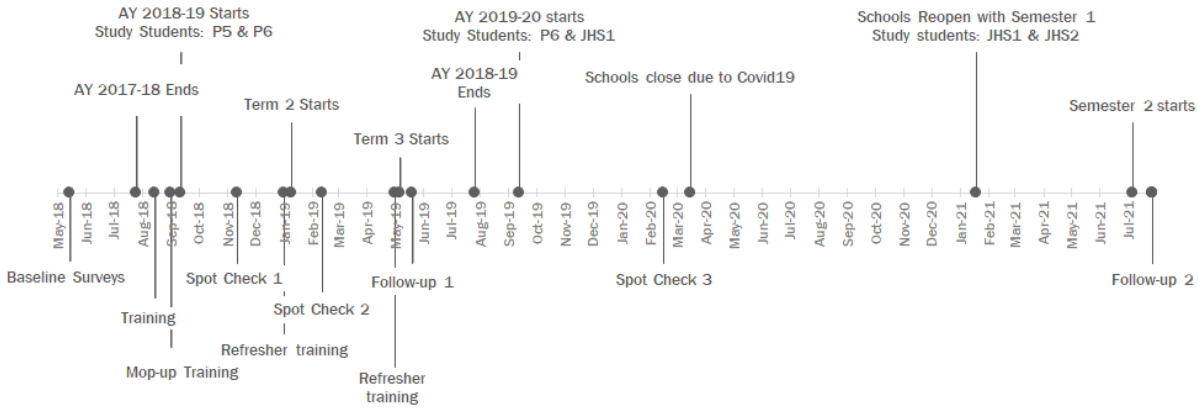
[Appendix Table A1-A23 about here]

Figure 1: Randomization Design



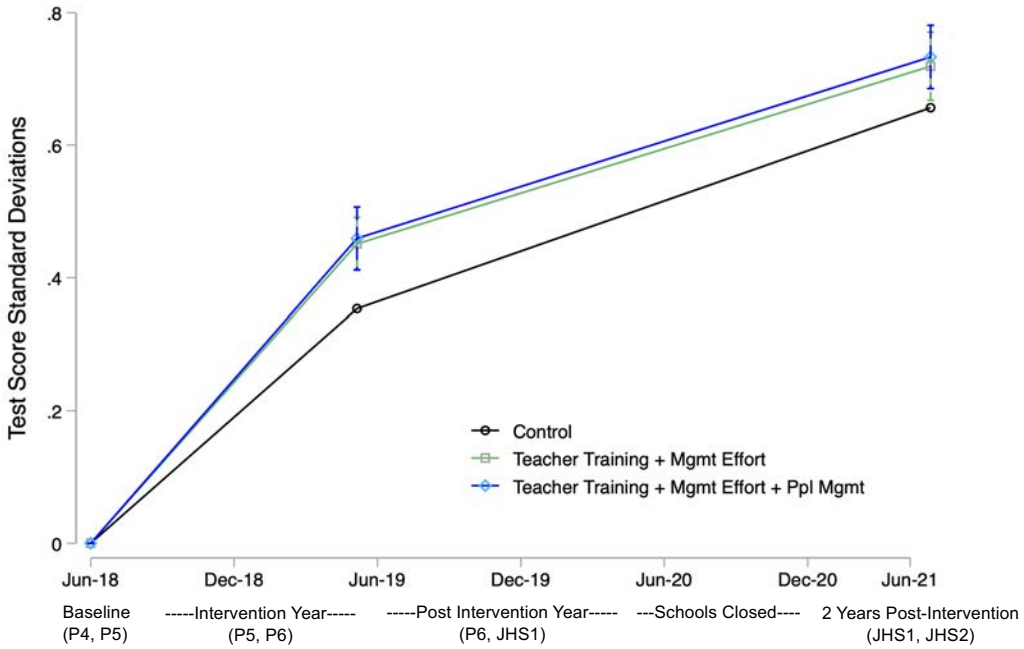
Notes: A circuit is a geographically proximate group of approximately 9 schools assigned to a single circuit supervisor. The study involved only 1 or 2 schools per circuit. Circuits were randomly assigned to the Management Effort arm or the Management Effort + People Management treatment. Within each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

Figure 2: Timeline of Study Activities



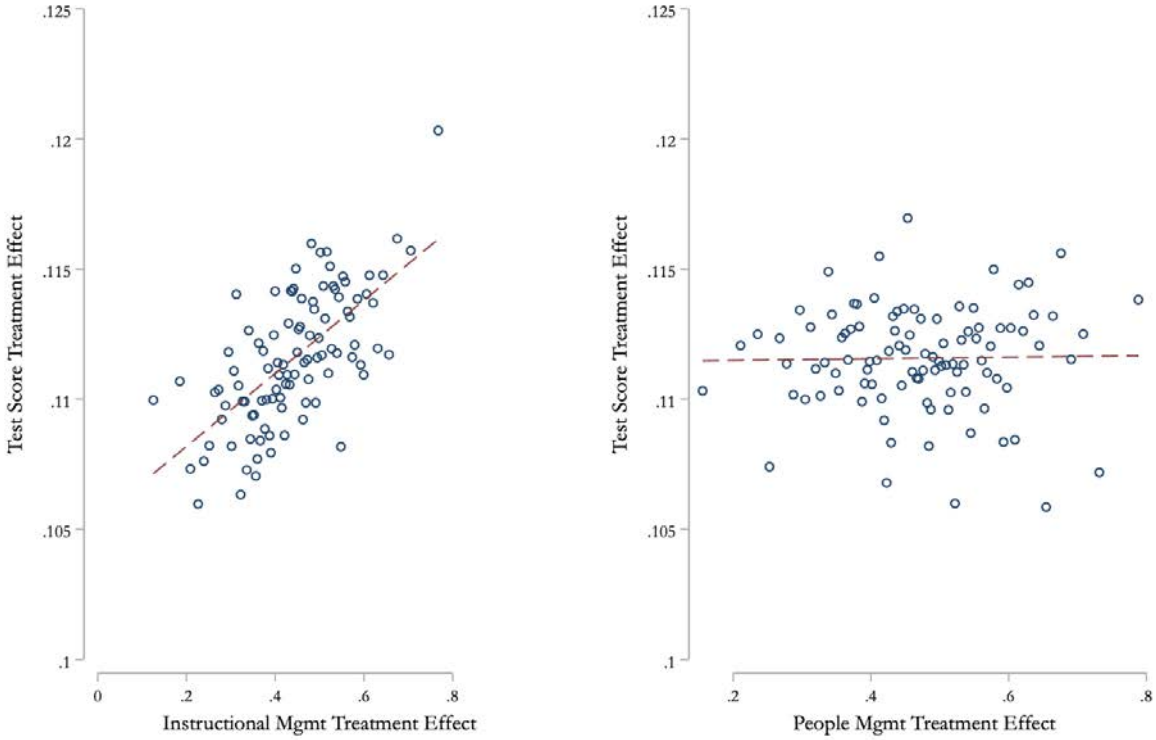
Notes: Activities above the timeline correspond to school calendar dates, while activities below the timeline correspond to intervention and data collection activities.

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' grades on the x-axis assumes timely progression.

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, Instructional, and People Management. 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 Instructional Management. The right graph shows the correlation between the treatment effects on test scores and People Management.

Table 1: Effects on Management Practices

	At Least One Mgr Obs 30 min	**Instructional Management Index**	**People Management Index**	**Other Management Index**
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Effort	0.357***	0.264*	0.228	0.367**
	(0.079)	(0.139)	(0.154)	(0.144)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.368***	0.346***	0.663***	0.641***
	(0.081)	(0.126)	(0.156)	(0.165)
P-Value Same Effect	0.88	0.40	0.01	0.07
Observations	210	210	210	210
R^2	0.21	0.23	0.34	0.21
Mean Dep., Control	0.39	-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 School Operations and Personnel Attendance at Unannounced Spot Checks

Attendance:	**School in Session** (1)	**Head Teachers** (2)	**Teachers** (3)
Teacher Training + Mgmt Effort	0.029*	0.111*	0.042
	(0.016)	(0.064)	(0.037)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 3: 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 Effort	0.130***	0.470***	0.572***	0.787***
	(0.032)	(0.060)	(0.047)	(0.018)
Teacher Training + Mgmt Effort + Ppl Mgmt	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. 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. Column 4: unit of observation is school-level.

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

Table 4: Effects on Norms of Behavior

	Teacher Opinions of Teacher Shirking (1)	**Teacher Opinions of All Roles Shirking** (2)	**HT Opinions of Teacher Shirking** (3)	**HT Opinions of All Roles Shirking** (4)
Teacher Training + Mgmt Effort	-0.261**	-0.310***	-0.251*	-0.038
	(0.108)	(0.105)	(0.152)	(0.153)
Teacher Training + Mgmt Effort + Ppl Mgmt	-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 each head teacher's responses.

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

Table 5: Effects on Student Test Scores

	Combined Score (1)	**Math Score** (2)	**English Score** (3)
Teacher Training + Mgmt Effort	0.108***	0.140***	0.065***
	(0.021)	(0.026)	(0.022)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 6: Management Outcomes in Year After Intervention

	Instructional Management	**People Management**	**HT Retained**	**Teacher Retained**
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Effort	0.138	0.455**	-0.010	0.206***
	(0.131)	(0.211)	(0.071)	(0.042)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.311**	0.856***	0.117*	0.092**
	(0.129)	(0.227)	(0.071)	(0.046)
P-Value Same Effect	0.12	0.11	0.07	0.01
Observations	208	208	208	687
R^2	0.32	0.32	0.26	0.16
Mean Dep., Control	-0.00	-0.00	0.70	0.49

Notes: 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 7: Teacher DI Implementation in Year After Intervention

Dependent Var:	**Class In Levels** (1)	**DI Imp Index** (2)	**Active Learning Index** (3)
Teacher Training + Mgmt Effort	0.416***	0.723***	0.047
	(0.072)	(0.038)	(0.135)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 spot check, 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. Linear probability models. Columns 1 and 2: Unit of observation is school. Column 3: Unit of observations is school-classroom observation. Regressions also 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. Column 1: Whether the class was split by levels during the spot check. Column 2: The percent of 5 different measures of differentiated instruction that the respondent reports doing; Control group was imputed to be zero. Column 3: An index of active learning variables.
$*p < 0.10, **p < 0.05, ***p < 0.01$

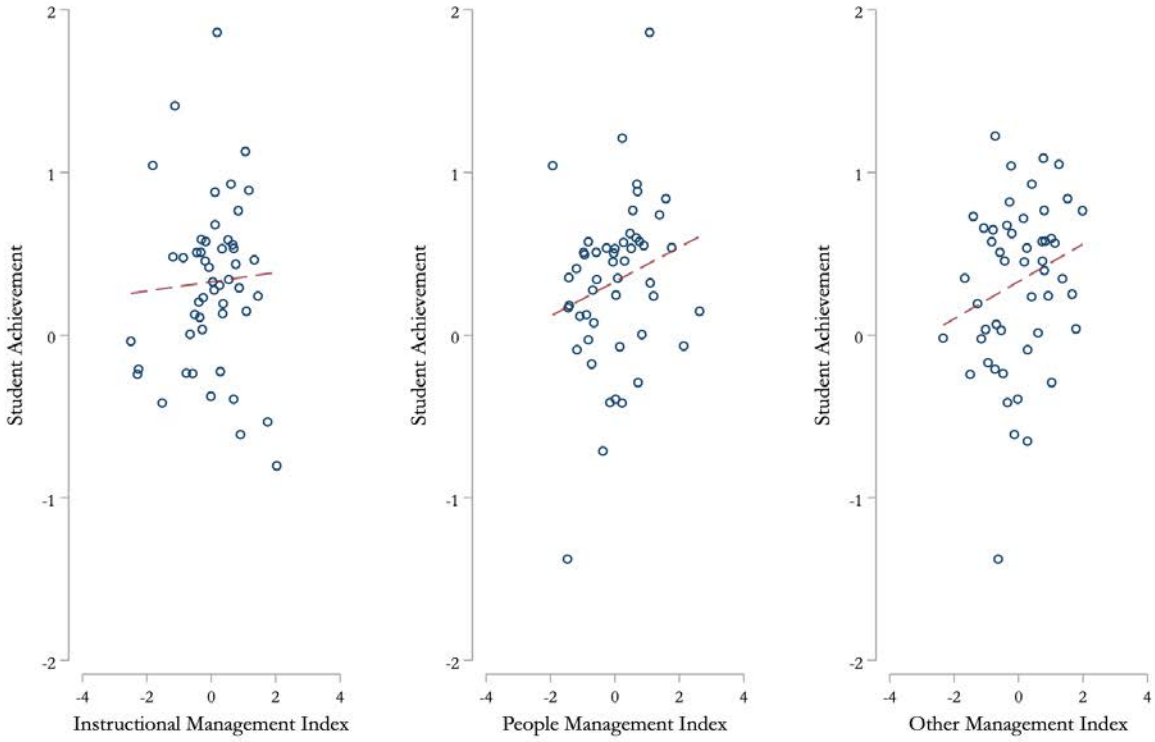
Table 8: Persistent Student Test Scores Effects

	Combined Score (1)	**Math Score** (2)	**English Score** (3)
Teacher Training + Mgmt Effort	0.062**	0.077**	0.055**
	(0.026)	(0.031)	(0.027)
Teacher Training + Mgmt Effort + Ppl Mgmt	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.
$*p < 0.10, **p < 0.05, ***p < 0.01$

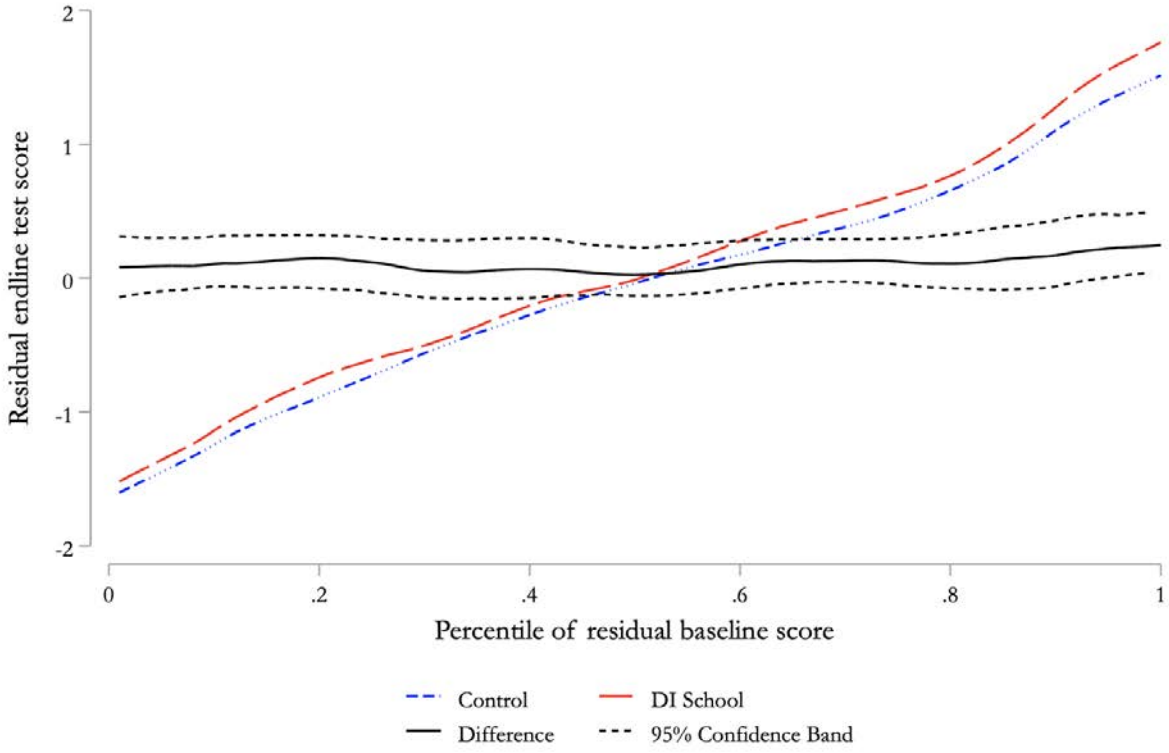
Appendix

Figure A1: Control Group Correlation Between Management and Student Test Scores



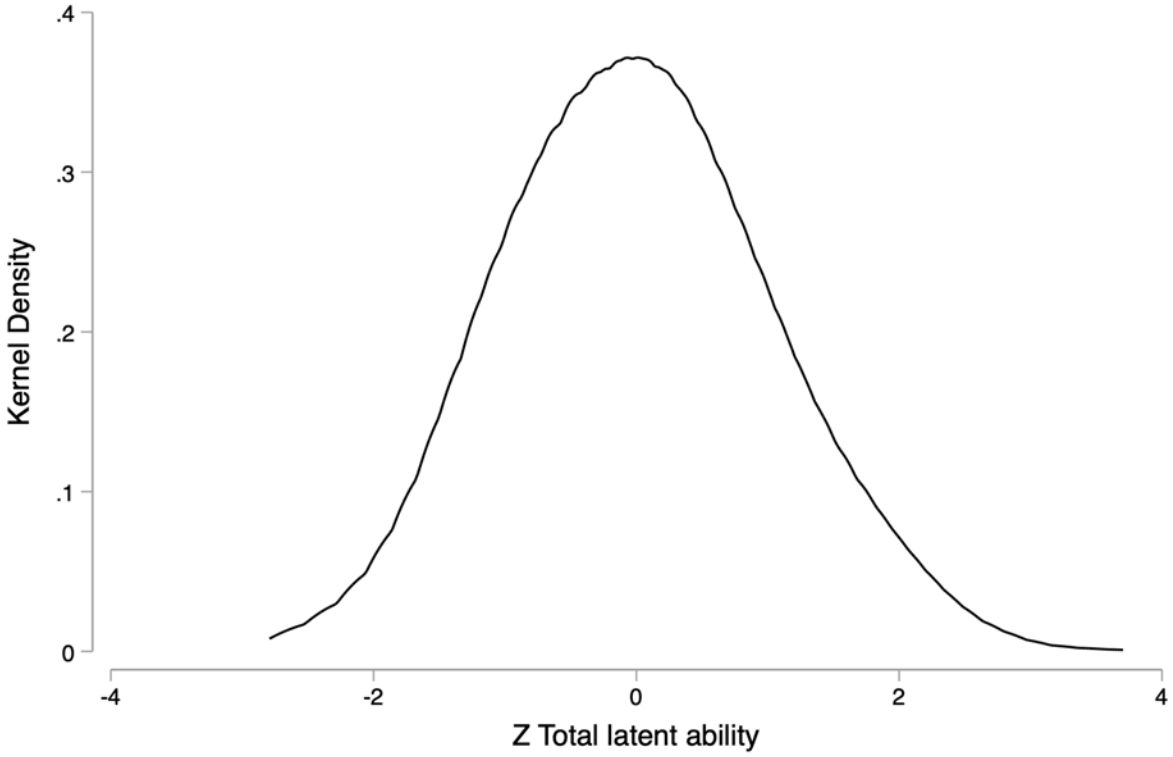
Notes: Plot of the binned scatterplot with the best-fit line added of the correlation between standardized student test scores (y-axis) and the management practice indices for the control group.

Figure A2: Non-Parametric Test Results



Notes: This plot shows the non-parametric test score effects of the student's combined score on math and English by pooled treatment status. 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 line shows the effects by baseline percentile distribution for the pooled treatment group, and the blue line shows the effects by baseline percentile for the control group. The difference between the pooled treatment and the control group is displayed in black. Dashed lines display the 95 percent confidence intervals.

Figure A3: 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.

Table A1: Effects on Individual Components of Instructional Management Index

	Teacher Training + Mgmt Effort (1)	**Teacher Training + Mgmt Effort + Ppl Mgmt** (2)	**P-value** (3)	**Control Mean** ' (4)
Average Number of P4-P6 Classroom Obsevations (5 min) by HT (HT)	0.265	-0.599	0.019	3.299
	(0.489)	(0.418)		
Pupils test scores are used to inform promotion (HT)	0.243**	0.215**	0.769	4.286
	(0.098)	(0.092)		
How often P4-P6 teachers use lesson plans (HT)	0.029	0.124	0.140	3.829
	(0.097)	(0.082)		
Number CS classroom observations (HT)	1.814***	2.862***	0.045	2.171
	(0.429)	(0.430)		
Teachers pay attention to indvidual student needs (HT)	0.129	0.043	0.302	4.114
	(0.109)	(0.113)		
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 Classroom Obsevations (5 min) in study schools (CS)	-0.000	1.098***	0.000	2.412
	(0.184)	(0.226)		
Number of CS Classroom Obsevations (30 min) in study schools (CS)	0.000	2.061***	0.000	3.432
	(0.355)	(0.449)		
CS Classroom Obsevations Frequency (5 min) (CS)	0.000	0.281**	0.043	2.985
	(0.132)	(0.138)		
CS Classroom Obsevations (30 min) Frequency (CS)	-0.000	0.088	0.527	2.909
	(0.125)	(0.140)		

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 Instructional 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 Effort. Column 2: Effect of the Teacher Training + Management Effort + People Management 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 A2: Effects on Individual Components of People Management Index

	Teacher Training + Mgmt Effort (1)	Teacher Training + Mgmt Effort + Ppl Mgmt (2)	P-value (3)	Control Mean (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 school visits by CS (HT)	1.643**	2.224***	0.459	5.971
	(0.695)	(0.721)		
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)		

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 A3: Effects on Individual Components of Other Management Index

	Teacher Training + Mgmt Effort	Teacher Training + Mgmt Effort + Ppl Mgmt	P-value	Control Mean '
	(1)	(2)	(3)	(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 Effort. Column 2: Effect of the Teacher Training + Management Effort + People Management 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: Summary Statistics

	—Control—	Tch Training+Mgmt Effort	Tch Training+Mgmt Effort+ Ppl Mgmt	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)	
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 A5: Management Outcomes Indices By Respondent Type

Index:	**Instructional Management**		**People Management**	
Reported by:	**Subordinate**	**Self**	**Subordinate**	**Self**
	(1)	(2)	(3)	(4)
Teacher Training + Mgmt Effort	0.598***	0.223	0.324**	0.078
	(0.132)	(0.146)	(0.161)	(0.140)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.574***	0.344**	0.657***	0.398**
	(0.123)	(0.134)	(0.150)	(0.154)
P-Value Same Effect	0.83	0.27	0.05	0.03
Observations	210	210	210	210
R^2	0.34	0.17	0.27	0.34
Mean Dep., Control	0.00	0.00	0.00	0.00

Notes: The outcome variable in all regressions are standardized 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 A6: 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 Effort	0.192***	0.206***	0.180***	0.470***
	(0.032)	(0.025)	(0.029)	(0.060)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 Effort	0.123***	0.148***	0.105***	0.370***
	(0.026)	(0.025)	(0.025)	(0.055)
Teacher Training + Mgmt Effort + Ppl Mgmt	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. Panel A uses all observations. Column 4 repeats Table 3, column 2. Panel B uses all observations conditional on a teacher being present during the entire observation.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A7: 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 Effort	0.144***	0.200***	0.201***	0.180***	0.473***
	(0.035)	(0.036)	(0.037)	(0.038)	(0.078)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 Effort	0.117***	0.185***	0.210***	0.180***	0.466***
	(0.038)	(0.040)	(0.038)	(0.039)	(0.080)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 A8: DI Implementation Index Components

Dependent Variable:	**Did Leveling Exam**	**Had DI Handbook**	**Imp DI Last 5 Days**	**Imp DI Today**	**Split in Levels**
	(1)	(2)	(3)	(4)	(5)
Teacher Training + Mgmt Effort	0.951***	0.740***	0.896***	0.779***	0.571***
	(0.015)	(0.038)	(0.023)	(0.031)	(0.047)
Teacher Training + Mgmt Effort + Ppl Mgmt	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. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table A9: Norms of Acceptable Behavior

	Teacher's Opinions of Students	**Teacher's Opinions of HTs**	**Teacher's Opinions of CS**	**HT Opinions of Students**	**HT Opinions of HTs**	**HT Opinions of CS**
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Training + Mgmt Effort	-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 Effort + Ppl Mgmt	-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 6. 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 A10: Work Stress and Burnout

	Teacher Stress & Burnout Index (1)	**HT Stress & Burnout Index** (2)
Teacher Training + Mgmt Effort	0.042	0.043
	(0.081)	(0.118)
Teacher Training + Mgmt Effort + Ppl Mgmt	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. Column 1: Teacher-reported. Column 2: Head teacher-reported. Standard errors are clustered at the school level.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table A11: Additional Test Score Specifications

	Combined Score (1)	**Math Score** (2)	**English Score** (3)
Panel A: Limited Covariates			
Teacher Training + Mgmt Effort	0.106***	0.134***	0.067**
	(0.033)	(0.036)	(0.033)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.081**	0.109***	0.050
	(0.037)	(0.039)	(0.037)
Panel B: Foundational Skills Only			
Teacher Training + Mgmt Effort	0.139***	0.183***	0.071***
	(0.027)	(0.034)	(0.022)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.169***	0.203***	0.104***
	(0.029)	(0.035)	(0.027)
Panel C: Upper-Level Items Only			
Teacher Training + Mgmt Effort	0.075***	0.086***	0.045*
	(0.023)	(0.027)	(0.026)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.054**	0.067**	0.033
	(0.026)	(0.030)	(0.027)
Panel D: Subset Asked of All Students			
Teacher Training + Mgmt Effort	0.118***	0.131***	0.075**
	(0.034)	(0.036)	(0.030)
Teacher Training + Mgmt Effort + Ppl Mgmt	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 A12: Test Scores

	(1)	(2)	(3)	(4)	(5)
Teacher Training + Mgmt Effort	0.106***	0.098***	0.098***	0.110***	0.106***
	(0.021)	(0.027)	(0.030)	(0.028)	(0.031)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.106***	0.091***	0.096***	0.123***	0.129***
	(0.024)	(0.029)	(0.028)	(0.033)	(0.035)
Teacher Training + Mgmt Effort X Baseline Test Score	0.011				
	(0.022)				
Teacher Training + Mgmt Effort + Ppl Mgmt X Baseline Test Score	0.002				
	(0.022)				
Teacher Training + Mgmt Effort X Female Student		0.023			
		(0.036)			
Teacher Training + Mgmt Effort + Ppl Mgmt X Female Student		0.034			
		(0.035)			
Teacher Training + Mgmt Effort X P4 Pupil at Baseline			0.020		
			(0.042)		
Teacher Training + Mgmt Effort + Ppl Mgmt X P4 Pupil at Baseline			0.023		
			(0.032)		
Teacher Training + Mgmt Effort X Either Parent Literate				-0.002	
				(0.034)	
Teacher Training + Mgmt Effort + Ppl Mgmt X Either Parent Literate				-0.029	
				(0.035)	
Teacher Training + Mgmt Effort X Above Median Amenities					0.002
					(0.043)
Teacher Training + Mgmt Effort + Ppl Mgmt 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 5. Column 5: Amenity index calculated at the school level.

Table A13: Student Tested at Endline

	Tested (1)	**Tested** (2)
Teacher Training + Mgmt Effort	-0.003	-0.003
	(0.006)	(0.006)
Teacher Training + Mgmt Effort + Ppl Mgmt	-0.015**	-0.015**
	(0.007)	(0.007)
Baseline Test Score	0.006	0.004
	(0.003)	(0.005)
Teacher Training + Mgmt Effort X Baseline Test Score		0.006
		(0.007)
Teacher Training + Mgmt Effort + Ppl Mgmt 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 A14: 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 Effort	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 Effort + Ppl Mgmt	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 5. See additional notes in Table 5.

Table A15: Student Attendance

	Student Absent From School (1)	**School Reports Student Not Enrolled** (2)	**Student Self-Reported Dropout** (3)
Teacher Training + Mgmt Effort	0.031**	0.028***	0.001
	(0.014)	(0.009)	(0.004)
Teacher Training + Mgmt Effort + Ppl Mgmt	-0.000	0.019**	0.004
	(0.013)	(0.009)	(0.005)
P-Value Same Effect	0.02	0.38	0.40
Observations	11,569	5,893	5,608
R^2	0.03	0.02	0.02
Mean Dep., Control	0.17	0.06	0.01

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

Table A16: Student Opinions About School

	Fave subject is Math or Eng (1)	**Like Math or English Very Much** (2)	**Want to Attend SHS** (3)
Teacher Training + Mgmt Effort	0.012	0.046***	-0.002
	(0.017)	(0.014)	(0.007)
Teacher Training + Mgmt Effort + Ppl Mgmt	-0.028	0.019	0.000
	(0.018)	(0.015)	(0.006)
P-Value Same Effect	0.02	0.05	0.72
Observations	5,608	5,608	5,550
R^2	0.06	0.05	0.03
Mean Dep., Control	0.79	0.79	0.97

The sample is all student available for the endline assessments. Column 1: The outcome variable is whether or not the child reported their favorite subject was math or English. Column 2: The outcome variable is whether or not the child reported liking math or English Column 3: Whether the child aspired to attend senior high school. Standard errors in parentheses, clustered at the school level. * p<0.10, ** p<0.05, *** p<0.01

Table A17: Management Outcomes in Year After Implementation: Individual Components of Instructional Mgmt Index

	Teacher Training + Mgmt Effort (1)	**Teacher Training + Mgmt Effort + Ppl Mgmt** (2)	**p-value** (3)	**Mean for Control '** (4)
Average Number of P4-P6 Classroom Obsevations (5 min) by HT (HT)	0.113	-0.739	0.100	3.951
	(0.628)	(0.488)		
Pupils test scores are used to inform promotion (HT)	0.286*	0.376***	0.501	4.043
	(0.147)	(0.142)		
How often P4-P6 teachers use lesson plans (HT)	-0.009	-0.091	0.422	3.836
	(0.087)	(0.104)		
Number CS classroom observations (HT)	-0.257	-0.458	0.472	2.188
	(0.267)	(0.285)		
Teachers pay attention to indvidual student needs (HT)	0.067	0.180*	0.269	4.058
	(0.114)	(0.103)		
HT Classroom Observations (5 min) (T)	-0.433	-0.513	0.841	4.167
	(0.572)	(0.493)		
CS Classroom Observations (5 min) (T)	0.131	0.129	0.982	0.924
	(0.092)	(0.094)		
HT Classroom Observation Feedback (T)	0.071*	0.109***	0.324	0.670
	(0.039)	(0.037)		
CS Classroom Observation Feedback (T)	0.082**	0.108***	0.524	0.437
	(0.040)	(0.042)		
Average Number of P4-P6 Classroom Obsevations (30 min) by HT (HT)	0.246	0.319	0.720	1.208
	(0.228)	(0.197)		
How often does HT observe teaching in school (HT)	0.134	0.084	0.706	3.406
	(0.132)	(0.145)		
HT Classroom Observations (30 min) (T)	0.134	0.126	0.957	1.351
	(0.266)	(0.236)		
CS Classroom Observations (30 min) (T)	0.148**	0.232***	0.221	0.383
	(0.062)	(0.062)		
Number of CS Classroom Obsevations (5 min) in study schools (CS)	0.012	0.160	0.402	1.504
	(0.152)	(0.180)		
Number of CS Classroom Obsevations (30 min) in study schools (CS)	-0.000	-0.259	0.211	1.145
	(0.180)	(0.210)		
CS Classroom Obsevations Frequency (5 min) (CS)	0.004	-0.214	0.148	3.076
	(0.131)	(0.150)		
CS Classroom Obsevations (30 min) Frequency (CS)	-0.001	0.027	0.850	2.606
	(0.125)	(0.149)		

Notes: 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 A18: Management Outcomes in Year After Implementation: Individual Components of People Mgmt Index

	Teacher Training + Mgmt Effort (1)	Teacher Training + Mgmt Effort + Ppl Mgmt (2)	p-value (3)	Mean for Control (4)
HT provides constructive feedback (HT)	-0.031	0.022	0.630	4.449
	(0.102)	(0.111)		
How often HT gives Ts suggestions about improving teaching (HT)	0.059	0.103	0.705	3.435
	(0.128)	(0.125)		
An important part of HT job is to ensure teaching skills are improving (HT)	0.134	0.288***	0.095	4.406
	(0.104)	(0.097)		
HT encourages teachers to try new practices (HT)	0.054	0.041	0.886	4.333
	(0.086)	(0.086)		
How often HT takes initiative to discuss matters with Ts (HT)	0.081	0.069	0.922	3.464
	(0.123)	(0.130)		
Number of staff meetings each term by HT (HT)	2.770**	4.865***	0.196	6.797
	(1.338)	(1.453)		
Number of school visits by CS (HT)	-0.118	-0.048	0.886	4.855
	(0.447)	(0.465)		
Number of HT meetings with CS (HT)	0.077	0.015	0.931	4.391
	(0.602)	(0.662)		
CS is a valuable mentor (HT)	-0.092	0.111	0.078	4.522
	(0.107)	(0.094)		
CS works with HT to solves problems (HT)	0.042	-0.029	0.531	4.420
	(0.108)	(0.121)		
CS performs valuable work for school (HT)	-0.095	-0.022	0.508	4.391
	(0.114)	(0.105)		
CS is a good manager (HT)	-0.066	-0.129	0.580	4.391
	(0.112)	(0.114)		
Number of staff meetings by HT (T)	0.003	-0.253	0.353	3.212
	(0.344)	(0.265)		
Number of staff meetings by CS (T)	-0.154	-0.078	0.553	1.610
	(0.166)	(0.164)		
CS provides useful feedback (T)	0.061	0.103**	0.270	0.406
	(0.039)	(0.042)		
HT provides useful feedback HT (T)	0.084**	0.124***	0.311	0.603
	(0.041)	(0.039)		
HT feedback mentions something T did well (T)	0.183**	0.253***	0.354	0.504
	(0.077)	(0.081)		
HT feedback offers sugg. for improvement (T)	0.061	0.111***	0.271	0.621
	(0.044)	(0.042)		
CS feedback offers sugg. for improvement (T)	0.142**	0.162**	0.813	0.277
	(0.066)	(0.077)		
CS feedback mentions something T did well (T)	0.169**	0.139*	0.730	0.284
	(0.071)	(0.076)		
T feel valued and appreciated (T)	0.040	0.035	0.912	0.367
	(0.050)	(0.050)		
HT has helped T become a better Teacher (T)	0.035	0.090*	0.302	0.295
	(0.050)	(0.052)		
CS has helped T become a better Teacher (T)	0.070	0.140***	0.139	0.200
	(0.047)	(0.047)		
HT is a good manager (T)	-0.042	-0.006	0.502	0.360
	(0.053)	(0.055)		
Number of CS meetings with each HT in their circuit (CS)	0.002	-0.578	0.194	4.127
	(0.419)	(0.448)		
Number of CS meetings with all staff their circuit (CS)	0.015	-0.460	0.229	2.476
	(0.386)	(0.393)		
Number of CS visits to schools their circuit (CS)	0.009	0.040	0.908	3.455
	(0.234)	(0.272)		
Number of CS meetings with all HTs their circuit (CS)	-0.056	1.066	0.398	1.455
	(0.475)	(1.300)		
Suggestions to teachers frequency (CS)	-0.006	0.007	0.920	3.485
	(0.113)	(0.136)		
Strongly agree CS valuable mentor for teachers (CS)	-0.002	-0.031	0.641	0.864
	(0.054)	(0.063)		
Strongly agree CS valuable mentor for headteachers (CS)	0.002	-0.009	0.853	0.864
	(0.052)	(0.060)		
Strongly agree CS provides constructive feedback to teachers (CS)	0.003	0.010	0.919	0.636
	(0.061)	(0.064)		
Strongly agree CS perform valuable work for school (CS)	0.001	0.094	0.281	0.600
	(0.085)	(0.087)		
CS provides suggestions to HT (CS)	0.002	0.077	0.280	4.773
	(0.069)	(0.069)		

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

	DI Implementation			
	Low Teacher Retention Sample (1)	High Teacher Retention Sample (2)	Not Retained HT Sample (3)	Retained HT Sample (4)
Teacher Training + Mgmt Effort	0.336	0.353***	0.445**	0.457***
	(0.204)	(0.090)	(0.199)	(0.087)
Teacher Training + Mgmt Effort + Ppl Mgmt	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: The outcome variable in each regression is the dummy variable indicating whether or not the school was still implementing differentiated instruction in the yer 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 A20: Student Test Scores Two Years After Implementation: Heterogeneous Treatment Effects

	(1)	(2)	(3)	(4)	(5)
Teacher Training + Mgmt Effort	0.063**	0.086**	0.063*	0.033	0.051
	(0.026)	(0.034)	(0.035)	(0.038)	(0.042)
Teacher Training + Mgmt Effort +Ppl Mgmt	0.079***	0.068**	0.057*	0.074*	0.084**
	(0.024)	(0.032)	(0.031)	(0.038)	(0.036)
Teacher Training + Mgmt Effort X Baseline Test Score	0.036				
	(0.028)				
Teacher Training + Mgmt Effort +Ppl Mgmt X Baseline Test Score	-0.005				
	(0.024)				
Teacher Training + Mgmt Effort X Female Student		-0.051			
		(0.044)			
Teacher Training + Mgmt Effort +Ppl Mgmt X Female Student		0.019			
		(0.041)			
Teacher Training + Mgmt Effort X P4 Pupil at Baseline			-0.001		
			(0.043)		
Teacher Training + Mgmt Effort +Ppl Mgmt X P4 Pupil at Baseline			0.044		
			(0.037)		
Teacher Training + Mgmt Effort X Either Parent Literate				0.048	
				(0.047)	
Teacher Training + Mgmt Effort +Ppl Mgmt X Either Parent Literate				0.005	
				(0.046)	
Teacher Training + Mgmt Effort X Above Median Amenities					0.021
					(0.057)
Teacher Training + Mgmt Effort +Ppl Mgmt 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 A21: 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 Effort	-0.017	0.003	0.054**	0.046***
	(0.013)	(0.011)	(0.023)	(0.015)
Teacher Training + Mgmt Effort +Ppl Mgmt	-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. 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. 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 A22: Student Tested at Longer-Term Follow-Up

	Tested (1)	**Tested** (2)
Teacher Training + Mgmt Effort	-0.023*	-0.023*
	(0.012)	(0.012)
Teacher Training + Mgmt Effort + Ppl Mgmt	-0.032**	-0.032**
	(0.013)	(0.013)
Teacher Training + Mgmt Effort X Baseline Test Score		0.007
		(0.012)
Teacher Training + Mgmt Effort + Ppl Mgmt 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 A23: 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 Effort	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 Effort +Ppl Mgmt	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 8. See additional notes in Table 8.

Table A24: Management Outcomes Indices (FU2 Only)

Index:	**Instructional Management** (1)	**People Management** (2)	**Teacher Retention** (3)	**HT Retention** (4)
Teacher Training + Mgmt Effort	0.086	0.112	0.208***	0.029
	(0.154)	(0.159)	(0.045)	(0.080)
Teacher Training + Mgmt Effort + Ppl Mgmt	0.051	0.106	0.058	0.051
	(0.156)	(0.171)	(0.045)	(0.087)
P-Value Same Effect	0.81	0.97	0.00	0.79
Observations	209	209	687	210
R^2	0.22	0.19	0.17	0.13
Mean Dep., Control	0.00	0.00	0.43	0.51

Notes: Col 1: The z-score of of instructional management variables in the second follow-up survey. Includes the baseline measure of the management index as a control variable. Col 2: The z-score of people management variables in the second follow-up survey. Includes the baseline measure of the management index as a control variable. Col 3: Whether or not teachers from the first spot check were still working in the school at the second follow-up survey. Regression also includes controls for age (and the square), sex, years of experience (and the square), and the teacher's class. Col 4: Whether or not the HT from baseline was still the HT of the school in the school follow-up survey. All regressions include district fixed effects and cluster standard errors at the school level.

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