Policy Research Working Paper 9535

School Management, Grants, and Test Scores

Experimental Evidence from Mexico

Mauricio Romero
Juan Bedoya
Monica Yanez-Pagans
Marcela Silveyra
Rafael de Hoyos

WORLD BANK GROUP
Education Global Practice
February 2021

Abstract

This paper presents the results of a large-scale randomized experiment conducted across 1,496 public primary schools in Mexico. The experiment identifies the impact on schools' managerial capacity and student test scores of providing schools with: (a) cash grants, (b) managerial training for school principals, or (c) both. The school principals' managerial training focused on improving principals' capacities to collect and use data to monitor students' basic numeracy and literacy skills and provide feedback to teachers on their instruction and pedagogical practices. After two years of implementing these interventions, the study finds that: (a) the cash grant had no impact on the student's test scores or the management capacity of school principals; (b) the managerial training improved school principals' managerial capacity but had no impact on students' test scores; and (c) the combination of cash grants and managerial training amplified the effect on the school principals' managerial capacity and had a positive but statistically insignificant impact on students' test scores.

This paper is a product of the Education Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at rdehoyos@worldbank.org.

School Management, Grants, and Test Scores: Experimental Evidence from Mexico*

Mauricio Romero[†] Juan Bedoya[‡] Monica Yanez-Pagans[§]
Marcela Silveyra[§] Rafael de Hoyos[†§¶]

February 2, 2021

*Corresponding author: Mauricio Romero (mtromero@itam.mx). This study was possible thanks to the support of the Secteraría de Educación Pública (SEP) of Mexico. We are especially indebted to Pedro Velasco, Griselda Olmos, Lorenzo Baladrón, Germán Cervantes, Javier Treviño, and all the staff at SEP's Directorate of Education Management. We are especially grateful to Raissa Ebner, Renata Lemos, and Daniela Scur for their collaboration in this project's early stages and subsequent discussions. Karina Gómez provided excellent research assistance. The views expressed here are those of the authors alone and do not necessarily reflect the World Bank's opinions. Romero gratefully acknowledges financial support from the Asociación Mexicana de Cultura, A.C. All errors are our own.

[†]ITAM

[‡]Universidad de Cantabria

[§]The World Bank

[¶]Xaber

1 Introduction

Schools are complex organizations that are often poorly managed. Across developed and developing countries, schools have worse management practices than hospitals and manufacturing firms (Bloom, Lemos, Sadun, Scur, & Van Reenen, 2014; Bloom, Lemos, Sadun, & Van Reenen, 2015). At least two factors can explain schools' low management capacity. First, public schools often lack budgetary and administrative autonomy and are unable to implement strategies to address their challenges. Second, school principals often lack management skills. In most developing countries, principals are chosen according to seniority: although they have years of classroom experience, they may have no relevant managerial skills.

We study the implementation of a large-scale strategy, called *Escuela al Centro*, designed by the Government of Mexico to strengthen school autonomy and improve school principals' managerial capacity. This strategy was implemented nationwide for three consecutive school years: 2015–16, 2016–17, and 2017–18. *Escuela al Centro* included two main components: school grants and school principals' managerial training.[1] The school grants component provided annual cash grants to schools of USD 5–USD 50 per student conditional on a school improvement plan approved by the school council.[2] Schools used these school grants mainly to purchase basic supplies (e.g., chalk, toiletries), educational materials (e.g., books, projectors), conduct small infrastructure repairs and pay for cleaning and maintenance. The school principals' managerial training component provided training on collecting and using data to monitor students' basic numeracy and literacy skills and providing feedback to teachers on their instruction and pedagogical practices. The managerial training was delivered to school principals under two modalities: (i) training provided under a "train the trainer" cascade model; and (ii) training sessions delivered by a team of professional trainers.

We randomly assigned 1,496 eligible public primary schools in Mexico to one of three groups: (1) PEC group, which received a school grant and school principal's managerial training using the cascade model ($n = 599$), (2) PEC Plus group, which received a school grant and school principal's managerial training delivered by a team of professional trainers ($n = 698$); and (3) control group, which received the school principals' managerial training using the cascade model ($n = 199$).

Our experimental design allows us to identify and measure the impact of the fol-

[1] The school grants were initially part of a program called *Programa Escuelas de Calidad* (PEC), which later transitioned into a component of the *Escuela al Centro* strategy.

[2] School councils (known as *Consejos Escolares de Participación Social*) are formed by parents, teachers, the school principal, and the school supervisor.

lowing components of the *Escuela al Centro* strategy: (i) school grants alone (PEC versus Control); (ii) using professional trainers to train school principals vis-a-vis the cascade model (PEC Plus versus PEC); and (iii) the combined effect of school grants and using professional trainers to train school principals (PEC Plus versus Control).

We look at two primary outcomes: school principals' managerial practices and students' learning. To measure school principals' managerial practices, we use the Development World Management Survey (DWMS) (Lemos & Scur, 2016), designed to measure the quality of school principals' managerial practices in developing countries. The DWMS measures different dimensions of school principals' managerial practices, including operations management, people management, target setting, and monitoring.

We collected data on school principals' managerial practices at baseline (in late 2015) and two years after the program was implemented (in early 2018). To measure students' learning, we use data from a nationwide standardized test (PLANEA).[3]

Our results show a statistically significant improvement of .12 (p-value 0.018) standard deviations (σ thereafter) in school principals' managerial capacities among PEC Plus schools versus PEC and of .26σ (p-value .0074) when PEC Plus schools are compared with the controls. The effects of PEC versus the controls—i.e., the isolated effect of the school grants—is positive (.16σ) but statistically insignificant (p-value .2).

However, none of the improvements in managerial capacities translate into statistically significant impacts on student learning. Students in PEC Plus schools have test scores that are 0.03σ (we can rule out an effect greater than 0.08σ at the 95% level) and 0.05σ (we can rule out an effect greater than 0.08σ at the 95% level) higher than their counterparts in PEC and control schools, respectively. The treatment effect of the school grants alone (i.e., PEC versus control schools) is negative (-0.08σ) and we can rule out an effect greater than 0.01σ at the 95% level.

In short, the training delivered directly by professional trainers to school principals had a positive impact on managerial practices, and school grants amplified these positive effects. However, school grants without a complementary intervention to help school principals make better use of the additional resources had no impact on managerial capacities. None of the interventions had a significant effect on learning. In addition, there is little evidence of heterogeneity in treatment effects by baseline school characteristics on both learning outcomes or management practices.

[3] *Plan Nacional para la Evaluación de los Aprendizajes* (PLANEA) is a nationwide standardized test designed by the Mexican Education Evaluation Institute that measures Math and Spanish learning outcomes in grades 6, 9, and 12. PLANEA is aligned with the Mexican curriculum and is applied to a sample of students in all schools in the country. In schools with fewer than 40 students in the grade assessed, every student is tested. In schools with more than 40 students, a random sample is tested.

The interventions' failure to improve learning outcomes could be related to the weak contemporary correlation between managerial practices and test scores in Mexico (as measured by the DWMS). Our baseline data show that an improvement of 1σ in managerial practices generates an increase of less than 0.1σ in test scores, a weaker correlation than Bloom, Lemos, et al. (2015) reported for several countries. The experimental treatment effects on test scores that we observe are aligned with the expected treatment effects, assuming the contemporary correlation at baseline is causal.

The cost-effectiveness analysis shows that investing in professional trainers to improve school principals' management capacities, as opposed to relying on the cascade model, is an efficacious policy in the absence of a strategy to strengthen schools' financial autonomy (i.e., school grants). However, even though the positive effects of professional trainers on managerial capacities are compounded by the school grants, combining training with grants might not be cost-effective due to the grants' high cost. Neither policy (professional trainers, and their combination with school grants) is cost effective to improve learning outcomes due to the null effect on student test scores.

We contribute to the literature and policy debate on improving school management in low- and middle-income countries in three ways. First, we contribute to research that explores the relationship between school management and student outcomes (World Bank, 2007). Recent evidence, mostly from developed countries, demonstrates that management practices are an important determinant of school effectiveness. Using data for 39 charter schools in the United States, Dobbie and Fryer (2013) show that traditional school inputs such as class size and teaching certifications cannot explain school effectiveness differences. By contrast, school management practices, such as providing feedback to teachers and using data to guide instruction, are a significant determinant of school effectiveness (Fryer, 2014). In line with Dobbie and Fryer (2013)'s findings, Bloom, Lemos, et al. (2015) administer the World Management Survey (WMS)—which assesses managerial practices—in six developed countries plus Brazil and India and document a positive and statistically significant correlation between managerial practices and student learning outcomes. Recently, Lemos, Muralidharan, and Scur (2021) found a positive correlation between managerial practices and student value added in India. More recently, an experimental evaluation in the United States found that a training program to improve school principals' managerial practices had a positive effect on student learning outcomes one year after implementation (Fryer, 2017). Our baseline data add to the evidence base on the correlation between school management and learning outcomes. Moreover, we provide some of the first experimental estimates of the effectiveness of a strategy to improve school principals' managerial capacity on management practices and student learning outcomes in a developing country. A closely related paper by Muralidharan and Singh (2020) showed that an attempt to improve management quality in Indian schools by inducing principals to adopt "best practices" had no impact on student outcomes. As in our setting, the accountability and incentive structure for principals did not change. While management capacity has an important impact on learning, our results show that providing training to school principals does not improve test scores in the short run without a larger impact on managerial practices. A potential explanation for lack of impact is that managerial practices take longer to improve student education outcomes (see de Hoyos, Ganimian, and Holland (2020)).

Second, we add to the evidence from low- and middle-income countries that increasing schools' resources has no impact on student learning outcomes on its own.[4] We find that school grants had a null effect on managerial practices and student test scores two years after the start of implementation.

Third, we contribute to the literature showing complementarities across policies designed to improve school effectiveness (e.g., Mbiti et al. (2019); Gilligan, Karachiwalla, Kasirye, Lucas, and Neal (2019); Brunetti et al. (2020)). We identify synergies between high-fidelity managerial training and school grants.

2 Context and intervention

2.1 Context

Mexico's primary education system (grades 1 to 6) includes more than 14 million students and 573,000 teachers distributed across roughly 100,000 schools. The system is highly decentralized, comprising 32 state-level education systems that follow a common national curriculum and general guidelines from the Federal Secretariat of Public Education (Federal SEP, from its acronym in Spanish). However, each of the State-level Secretariats of Public Education is entirely administered by local governments. Unlike other countries in Latin America, the private education sector in Mexico is small — public schools account for 90% of the total primary enrollment (Elacqua, Iribarren, & Santos, 2018). There are also three types of public primary schools in the country: general primary schools (which account for most of the enrollment) and indigenous and

[4]For example, Glewwe, Kremer, and Moulin (2009) in Kenya, Blimpo, Evans, and Lahire (2015) in The Gambia, Das et al. (2013) in India, Pradhan et al. (2014) in Indonesia, Sabarwal, Evans, and Marshak (2014) in Sierra Leone, and Mbiti et al. (2019) in Tanzania. In the study that is perhaps most closely related to our own, experimental evidence from the Mexican state of Colima shows that school grants had no impact on student learning outcomes one year after the intervention (Garcia-Moreno, Gertler, & Patrinos, 2019).

community schools, which serve roughly 800,000 and 400,000 students, respectively.

Access to primary education in Mexico is high, with over 98% of children aged 6 to 12 being enrolled in the education system (World Bank, 2017b; Dirección General de Planeación, Programación y Estadística Educativa, 2018). However, the quality of education is low. Although almost all children graduate from primary school (World Bank, 2017a), fewer than half of them achieve a basic proficiency level in math and Spanish according to the latest national students' standardized tests (Instituto Nacional para la Evaluación de la Educación, 2018). Poor learning outcomes are even starker in schools located in marginalized areas, where only one in three students achieves a basic level of proficiency when graduating from primary school.

Mexico has many small, multi-grade schools that serve a low percentage of the student population. The smallest 40% of primary schools in the country serve 8.5% of the country's primary school students.[5] The existence of a large number of small schools across the country increases the governance challenges and calls for the development of different school-based management models adapted to small, multi-grade, and other types of schools.

These governance challenges are compounded by the high rotation of teachers and school principals and the lack of a system to regulate the entry and promotion of qualified teachers. For decades, the national teachers' union has influenced teachers' (and school principals') appointments (Álvarez, García-Moreno, & Patrinos, 2007). In 2013, the central government implemented a major education reform that defined and regulated the process used to hire and promote teachers and school principals. Using a new merit-based mechanism to appoint school principals, the government designed the *Escuela al Centro* strategy to enhance school autonomy and school principals' managerial capacities, with the ultimate objective of improving students' learning outcomes.

2.2 The *Escuela al Centro* strategy

The government implemented the *Escuela al Centro* strategy nationwide for three consecutive school years: 2015–16, 2016–17, and 2017–18. *Escuela al Centro* was designed around two main components: the provision of school grants and school principals' managerial training.[6]

[5]By comparison, the US public elementary school system has 67,000 schools serving 35 million students. Mexico has less than half of the student population of the United States, but it has 50% more schools.

[6]The description of the *Escuela al Centro* strategy is available at: http://www.dof.gob.mx/nota_detalle_popup.php?codigo=5488338, and the operating rules are available at: http://www.dof.gob.mx/nota_detalle.php?codigo=5509544&fecha=29/12/2017.

The school grant component consisted of a cash grant to schools provided annually conditional to a school improvement plan approved by the school council. The grants ranged from USD 1,500–15,000 depending on the school's size (about USD 5–50 per student). Schools used these grants to implement their annual school improvement plan and pay for basic school supplies and repairs.

The school principals' managerial training component focused on improving school principals' capacity to collect and use data to monitor students' basic numeracy and literacy skills and provide feedback to teachers on their instruction and pedagogical practices. To implement this training, the Federal SEP developed two tools: (i) a student assessment to monitor foundational students' skills (*Sistema de Alerta Temprana en Escuelas de Educación Básica*, SisAT) and (ii) a Stallings classroom observation tool to provide feedback to teachers on how to improve their instructional and pedagogical practices. The classroom observation tool was developed based on evidence that using school principals as teachers' coaches has a positive impact on student learning in Mexico(Secretaría de Educación Pública & Banco Internacional de Reconstrucción y Fomento, 2015).

SisAT is a student assessment designed to measure students' basic numeracy and literacy skills to identify lagging students and trigger early remedial actions. Teachers administer the SisAT assessment, which includes items from past national standardized student assessments. The teacher then inputs the results into a simple software that compiles and analyzes them. SisAt produces a detailed report for teachers and school principals to identify students with significant learning gaps. Moreover, the SisAT also provides information on what areas of the national curriculum are the most challenging for students and classrooms. de Hoyos, García-Moreno, and Patrinos (2017) and de Hoyos, Ganimian, and Holland (2019) show positive effects of providing school directors information on what areas of the national curriculum are the most challenging ones for students based on the national standardized learning assessments. Schools were free to decide when to apply the SisAT assessments to their students. However, most of them applied them at the beginning of the school year to generate baseline measures to include in their school improvement plans. As the school year progressed, schools applied the SisAT several times at their discretion to monitor the progress of their students' learning.

The Stallings classroom observation collects information on the teacher's use of time in the classroom, including the activities conducted, pedagogical practices, use of educational material, and level of students' engagement.(Stallings, 1977; Stallings & Molhlman, 1988) The application of this tool allowed school principals to collect data systematically to provide feedback to teachers on how to improve their instructional and pedagogical practices.

The Federal SEP developed high-quality training materials for both the SisAT and the Stallings classroom observation tool. The training for each of the tools consisted of 40 hours of instruction.[7] However, as *Escuela al Centro* was designed to have national coverage, the government designed a delivery model that allowed to progressively scale-up the delivery of the training for both tools. To do so, the Federal SEP established a cascade training model. Under this cascade training model, state-level education authorities selected 10% of all primary school supervisors to receive training on the SisAT and Stallings classroom observation tool from a professional trainers team. This team of professional trainers included Federal SEP's staff involved in the design of the two school principals' managerial tools. Subsequently, the trained supervisors were responsible for training the rest of the supervisors in their states. After all supervisors in a state were trained (either directly by the team of professional trainers or by their peers), they were responsible for training the school principals under their school jurisdictions.

The target population for *Escuela al Centro* were all public primary schools in the country, and the school principals' managerial training component was compulsory for all of these schools. In contrast, the school grant component was not available to all schools. In coordination with state-level education authorities, the Federal SEP issued a call for applications for schools to benefit from the school grant component. Among all public primary schools not participating in other programs to strengthen school autonomy that expressed interest in benefiting from the school grant component, schools located in disadvantaged areas were prioritized based on the government's marginalization index.[8] Overall, 53,413 public primary schools with a student population of 10,067,403 (69.12% of the students enrolled in primary schools in Mexico) benefited from the school grant component in the 2015–16 school year.

3 Research design and data

3.1 Interventions and implementation

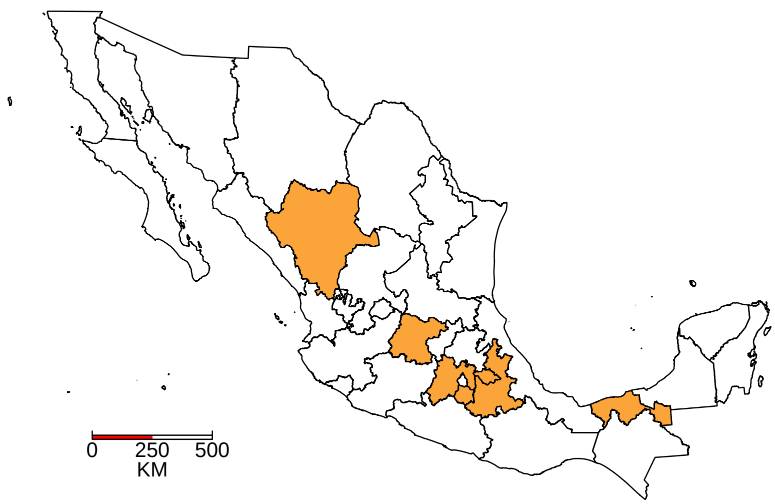
To test the effectiveness of *Escuela al Centro*, SEP invited all 32 states in Mexico to participate in this impact evaluation. From this nation-wide invitation, seven states were selected to be part of this research study: Durango, Estado de México, Morelos, Tlax-

[7] All training materials developed are available at the *Escuela al Centro* website: https://escuelaalcentro.com/intervenciones/descarga-los-materiales/.

[8] For details on the construction of the marginalization index, see: http://www.conapo.gob.mx/es/CONAPO/Indices_de_Marginacion_Publicaciones.

cala, Guanajuato, Tabasco, and Puebla.[9] The seven participating states' geographical distribution is shown in Figure 1.

Figure 1: States participating in the impact evaluation



Note: Geographical information on the administrative areas of Mexico comes from INEGI (2018). Figure A.1 provides the distribution of schools within each state.

As described above, *Escuela al Centro* was designed around two main components: the provision of school grants and school principals' managerial training. The managerial training component of *Escuela al Centro* was designed to progressively reach all school principals in public primary schools in the country using the "train the trainer" cascade model. However, as part of this impact evaluation, SEP was interested in testing the effectiveness of the cascade model versus professional training and it therefore agreed on providing training to school principals in selected schools directly from the team of professional trainers. Therefore, in the seven participating states, the managerial training component of *Escuela al Centro* was delivered to schools using two modalities: (i) training

[9] From the 32 states in Mexico, 14 states expressed interest in participating in the impact evaluation. However, only seven complied with the Federal SEP's required paperwork and were selected to participate in the impact evaluation.

provided by supervisors trained by a team of professional trainers, under a "train the trainer" cascade model (PEC); and (ii) training delivered directly to school principals by the team of professional trainers (PEC Plus).

In each of the seven states that participated in the impact evaluation, local education authorities invited all public primary schools to apply for the school grant component of *Escuela al Centro*. Among all schools that applied to the school grant component, we randomly assigned 1,496 schools to one of three groups: (1) PEC group, which received a school grant and school principals' managerial training using the cascade model ($n = 599$), (2) PEC Plus group, which received a school grant and school principals' managerial training delivered by professional trainers ($n = 698$); and (3) control group, which received school principals' managerial training using the cascade model ($n = 199$).[10] Table 1 summarizes the interventions received by the three groups included in our research design.

The design of this impact evaluation allows us to identify the causal effects on outcomes of the following components of the *Escuela al Centro* strategy: (i) school grants alone (PEC versus Control); (ii) using of professional trainers to train school principals vis-a-vis the cascade model (PEC Plus versus PEC); and (iii) the combined effect of school grants and using professional trainers to train school principals (PEC Plus versus Control).[11]

[10]Some of the school principals from the PEC Plus group also benefited from short-term leadership certificate training programs offered by the different state-level education authorities. These certificate programs focused on leadership issues, aligned with the national school principal's profile standards. As explained in more detail in Section 3.3, leadership practices are not used to assess school principals' managerial practices in the DWMS the instrument used in this paper to measure overall school principals' managerial practices. Further details of the short-term certification programs offered by the states are provided in Appendix A.2.

[11]While it is not possible to experimentally identify the impact of the school principals' managerial training delivered using the cascade model vis-a-vis no training at all, there is evidence that cascade training models tend to be relatively ineffective (Popova, Evans, Breeding, & Arancibia, 2018).

Table 1: Summary of the three experimental groups

	PEC Plus	PEC	Control
School budgetary autonomy			
School cash grant	Yes	Yes	
School principal managerial training			
Classroom observation + training by professionals	Yes		
Classroom observation + training via the cascade		Yes	Yes
Foundational skills measurement + training by professionals	Yes		
Foundational skills measurement + training via the cascade		Yes	Yes

3.2 Sampling and randomization

Our sampling frame included public primary schools with more than 60 students and excluded multi-grade schools (i.e., those with at least one classroom that includes students from different grades).[12] Therefore, public primary schools included in the experiment are larger (i.e., have more students and teachers) and more likely to be urban than the average public primary school in Mexico (see Table A.1).

The sampling strategy and final criteria for selecting schools to participate in the experiment varied slightly across the seven participating states.[13] Estado de México and Puebla had all three types of schools (PEC, PEC Plus and control), while the other five states had only PEC and PEC Plus schools. In Estado de México, due to its large population of eligible schools, local authorities decided to have two experiments. In one experiment, schools we randomly assigned to either the PEC or PEC Plus intervention.[14] In the second experiment, we sampled of 200 schools not participating in the first experiment from the metropolitan area of Mexico City and randomly assigned them to either PEC Plus or the control.[15]

In Puebla, schools were stratified based on the marginalization of their locality (high/low) and whether they were urban or rural. A random sample of 300 primary schools—

[12]Small schools were excluded as the managerial intervention was focused on training school principles to coach teachers. In small schools, school principals also take the role of teachers so they call for different school management models.

[13]For a detailed explanation of the sampling strategy, see Table A.3.

[14]Schools in the first experiment are not a representative sample of schools in the state. They are larger than average, more likely to be in rural areas of the state (see Figure A.1c), and have below-average achievement levels.

[15]Since all eligible schools in the first call for applications had already notified that they were selected to participate in the program, to have a control group, state-level education authorities issued a second call for applications. These schools are mainly in urban areas of Estado de México (see Figure A.1d) in what is considered the greater metropolitan area of Mexico City.

proportional to the group's size– was selected to participate in the experiment. We ranked schools within each group based on their enrollment. We assigned schools in a repeating sequence to PEC, PEC Plus, or the control; the order in which the sequence began was randomized.

In Morelos, Tlaxcala, Guanajuato, Tabasco, and Durango eligible public primary schools were randomly assigned to either PEC or PEC Plus. The sample sizes were 130, 110, 200, 165, and 200 schools, respectively.

3.3 Data

The data used in this study come from multiple sources. We collected primary data on the school principals' managerial practices and perceptions of the quality of the training they received. We also use secondary data from administrative records provided by SEP that include: (i) student learning outcomes; (ii) school marginalization index; and (iii) information on schools' infrastructure, enrollment rates, and the number of teachers. We have primary and secondary data for the school years 2015–16 (baseline) and 2017–18 (the follow-up). The baseline and follow-up coincide with the nationwide standardized test application dates, which allows us to measure the intervention's impact on both management practices and student test scores.

Information on schools' managerial practices was collected using the DWMS[16]—an adaptation of the WMS, originally developed to measure the quality of management practices in manufacturing firms in developed (Bloom & Van Reenen, 2007) and developing countries (Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013). The survey was subsequently adapted to measure management quality in the education and health sectors (Bloom, Lemos, et al., 2015; Bloom, Propper, Seiler, & Van Reenen, 2015). The WMS and DWMS are fully comparable; the latter can better identify granular differences in management practices at the lower end of the management quality distribution, where most public schools and hospitals in developing countries are located.

The DWMS adaptation to measure management practices in schools in developing countries that we use for this study consists of a recorded interview with the school principal. The interview includes 23 open-ended questions that collect information on four dimensions: operations management, people management, target setting, and monitoring.[17] The interviews were conducted by a team of two trained enumerators (one coder

[16]For more on the DWMS survey instrument, see Lemos and Scur (2016) and https://developingmanagement.org/

[17]In the case of Mexico, the adaptation of the DWMS included an additional dimension labeled leadership. Having this additional dimension responded to the government's need to better align the DWMS

and one conducting the interview) and lasted around two hours. While the DWMS is designed to be less subjective than the WMS to overcome the lower capacity of enumerators in developing countries, there is still considerable room for enumerator subjectivity in data coding. To ensure comparability over time, we assigned the same team of trained enumerators to code the audio files from all the original interviews. To ensure comparability across schools, we randomly assigned audio files to enumerators and control for enumerator fixed effects in all the regressions.[18] We conducted the baseline surveys of the DWMS between October 2015 and May 2016, and the follow-up from January to May 2018.[19]

School principals also completed two online surveys to assess the quality of implementing the managerial training vis-a-vis its design—one for the SisAT tool and one for the Stallings classroom observation tool. The surveys included questions about different elements of the tools and their associated training. Filling out the survey was not mandatory. As a result, a considerable number of school principals did not provide this information. Schools that answered the online surveys are statistically different from those that did not in almost every observable characteristic (see Tables A.13 and A.14). For completeness, we report some basic statistics from these two online surveys. However, the information provided by these surveys is not representative of our experimental sample, due to sample selection (i.e., it has differential attrition across treatments and within each treatment) and it is therefore not part of our main analysis.

Student learning outcomes are based on annual, student-level PLANEA test scores. PLANEA 6th grade was administered in June 2015 and June 2018. For this study, SEP provided access to anonymized student-level data for both years for all schools that participated in the evaluation.

In addition to test scores, PLANEA collects information on the location of each school. We use this information to match each school to the marginalization index of its locality. The marginalization index, estimated by the *Consejo Nacional de Población* (CONAPO), considers localities' deficiencies in terms of education, housing, population, and household income.

instrument to the rules of operation of *Escuela al Centro*. All the analysis reported in this paper excludes the leadership dimension when constructing the overall DMWS index for the sake of comparability with other settings.

[18]Unfortunately, 32% of audio files from the baseline, and 16% from the follow-up were damaged by the time we asked the enumerators to code the interviews. Schools with and without misplaced audio files in the endline are statistically indistinguishable in observable characteristics (see Table A.2). Thus, it is unlikely that the results are driven by differences in observable or unobservable characteristics between schools with and without functioning audio files.

[19]https://escuelaalcentro.com/ has a detailed timeline of when different rounds of data collection took place in each state.

Lastly, we use administrative school census data collected by the federal and state-level education authorities known as "*Formato 911*." Since 1998, *Formato 911* has been collected at the beginning and end of each school year. It gathers basic information on the number of students, the number of teachers and their qualifications, characteristics of the school principal, number of classrooms, and the school's geographic location. Using a unique school identifier (*Clave de Centro de Trabajo*), it is possible to match this school census data with the results from PLANEA.

3.4 Balance and attrition

Most student and school characteristics are balanced across treatment arms at baseline (see Table 2). The average school in our sample has 298 students, 9.9 teachers, and a pupil-teacher ratio of 29. 43% of schools are in rural areas and 35% are in areas categorized as poor or very poor by the government. The last two rows of the table show the fraction of schools for which we have DWMS data at the endline (in 2018) and PLANEA data in 2018. We have PLANEA data for nearly all schools ($\sim$99%) and DWMS data for $\sim$77% of schools. The proportion of schools with both PLANEA and DWMS data is balanced across treatments.

Table 2: Balance across treatment groups

	(1)	(2)	(3)	(4)	(5)	(6)
		Mean			**Difference**	
Variable	Control (C)	PEC (P)	PEC Plus (PP)	C-P	PP-P	PP-C
Students in math achievement L-IV (%)	9.38	7.77	8.36	-0.22	0.50	0.72
	(11.21)	(11.10)	(11.83)	(1.18)	(0.64)	(1.15)
Students in math achievement L-I (%)	55.23	60.05	59.61	1.40	0.33	-1.06
	(20.14)	(21.83)	(21.75)	(2.14)	(1.20)	(1.98)
Students in language achievement L-IV (%)	3.64	2.67	3.42	-0.14	0.60^{**}	0.74
	(4.76)	(3.86)	(6.15)	(0.51)	(0.28)	(0.55)
Students in language achievement L-I (%)	46.07	52.20	50.17	2.06	-0.33	-2.39
	(20.16)	(20.25)	(20.39)	(2.03)	(1.13)	(1.87)
Marginalization	0.36	0.38	0.32	0.01	-0.00	-0.01
	(0.48)	(0.49)	(0.47)	(0.01)	(0.02)	(0.01)
Urbanization	0.44	0.41	0.43	-0.01	-0.02	-0.01
	(0.50)	(0.49)	(0.50)	(0.04)	(0.02)	(0.04)
Number of students	354.92	272.75	303.55	11.85	12.12	0.27
	(210.53)	(163.64)	(183.76)	(18.76)	(8.84)	(18.53)
Number of teachers	11.15	9.27	10.05	0.33	0.31	-0.02
	(5.35)	(4.22)	(4.79)	(0.47)	(0.24)	(0.46)
Student-teacher ratio	30.75	28.35	29.20	0.77	0.53	-0.24
	(6.73)	(6.92)	(7.13)	(0.65)	(0.35)	(0.61)
DWMS endline missing	0.26	0.22	0.24	0.02	0.01	-0.01
	(0.44)	(0.41)	(0.43)	(0.04)	(0.02)	(0.04)
PLANEA endline missing	0.01	0.01	0.01	0.01	0.00	-0.01
	(0.07)	(0.08)	(0.08)	(0.01)	(0.00)	(0.01)
Observations	199	599	698	200	1,197	398

This table presents the means and standard deviations (in parentheses) for Control schools (Column 1), PEC schools (Column 2), and PEC Plus schools (Column 3). The differences, taking into account the randomization design (i.e., including strata fixed effects) between groups are in Columns 4–6, and standard errors (in parentheses) are clustered at the school level. Achievement level (L) refers to one of four knowledge domains granted to students in the results of the PLANEA 2015 exam. L-I refers to the lowest level, while L-IV refers to the highest level. *Marginalization* is a variable coded 1 for areas with "high" or "very high" marginalization, and 0 otherwise according to CONAPO. *Urbanization* is a variable coded 1 for schools located in an urban area, and 0 otherwise. The number of students and teachers is taken from Formato 911 for the 2015–2016 academic year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.5 Compliance

To measure compliance with the evaluation's original design, we compiled information on whether the schools' grants were delivered to PEC and PEC Plus schools (and not to the controls), and whether school principals in PEC Plus schools attended the training sessions on the use of the Stallings classroom observation tool and the foundational skills monitoring software (SisAT). However, the training information comes from self-reported data. Since the characteristics of schools that answered the survey are different from those that did not (see Tables A.13 and A.14), these results should be interpreted with caution. In addition, the training that principals in PEC and Control schools report receiving is through the cascade model, while the training that the PEC Plus principals report was imparted by professional trainers. Thus, the training hours are not directly comparable. Because of the sample selection bias in the compliance measures and the different nature of the "training hours" across treatment arms, we do not attempt to estimate a local average treatment effect (LATE), using the treatment assignment as an instrument for the training hours principals report.

According to administrative data from SEP, every PEC and PEC Plus school received a grant, while no control school did. While virtually no principals in the control or PEC groups completed the full training on the use of either the Stallings tool or the SisAT, some of them received some training (10–39 hours) through the cascade model. Approximately 30% of principals in control schools received some training, compared to roughly 40% of those in PEC schools (see Columns 1 and 2 of Table 3). About one-quarter of principals in PEC Plus schools (20–25%) completed the tools' training and roughly 80% received some training by professional trainers. This difference between PEC Plus and either PEC or Control schools is statistically significant for both completed training and for the indicator for some training.

Table 3: Compliance across treatment groups

	(1)	(2)	(3)	(4)	(5)	(6)
		Mean			Difference	
Variable	Control (C)	PEC (P)	PEC Plus (PP)	C-P	PP-P	PP-C
Panel A: Stallings classroom observation tool						
All training sessions (40 hours)	0.00	0.01	0.25	-0.27	0.19***	0.46***
	(0.00)	(0.12)	(0.43)	(0.05)	(0.02)	(0.05)
Some training sessions (10-40 hours)	0.26	0.42	0.84	-0.28	0.39***	0.67***
	(0.44)	(0.49)	(0.36)	(0.06)	(0.03)	(0.06)
Observations	80	361	703	13	970	264
Panel B: Foundational skills measurement tool (SisAT)						
All training sessions (40 hours)	0.01	0.01	0.20	-0.06	0.17***	0.23***
	(0.08)	(0.08)	(0.40)	(0.03)	(0.01)	(0.03)
Some training sessions (10-40 hours)	0.29	0.38	0.74	-0.17	0.35***	0.52***
	(0.46)	(0.49)	(0.44)	(0.05)	(0.03)	(0.04)
Observations	169	466	623	153	996	352

This table presents the means and standard deviations (in parentheses) for Control schools (Column 1), PEC schools (Column 2), and PEC Plus schools (Column 3). The differences, taking into account the randomization design—i.e., including strata fixed effects—between groups are in Columns 4–6, and standard errors (in parentheses) are clustered at the school level. Panel A has information on whether the school principal attended the training sessions for the Stallings classroom observation tool (and how many hours). Panel B has information on whether the school principal attended the training sessions on SisAT (and how many hours). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4 Results

4.1 Correlation between management (DWMS) and learning

We first explore the correlation between learning outcomes and DWMS at baseline. The goal is to replicate the analysis in Bloom, Lemos, et al. (2015) and compare our results with those previously found in the literature on the magnitude of the relationship between student learning outcomes and school management measured by the DWMS.

In our data, better management quality, as measured by the DWMS, is only marginally correlated with better educational outcomes (see Table 4). A one-standard-deviation increase in the DWMS index is associated with an increase of 0.00σ–0.02σ in student test scores. We follow Bloom, Lemos, et al. (2015) and control for the number of pupils in the school, the pupil/teacher ratio, and the marginalization index (Column 4). We also control for measurement error by adding interviewer fixed effects (Column 5). The point estimate is robust to various controls, and is never statistically significant. By comparison, Bloom, Lemos, et al. (2015) find that a one-standard-deviation increase in the WMS index is associated with an increase in pupil outcomes of 0.2–0.4σ. In Brazil, the setting included in their study closest to Mexico, a one-standard-deviation increase in the WMS index is associated with an increase in pupil outcomes of 0.1σ. Thus overall, we find a lower correlation between outcomes and management than previously documented in other countries.

Of the four components of the DWMS (operations, monitoring, targets, and people), targets is the most correlated with student outcomes, followed by monitoring and people; none of them show a statistically significant correlation with test scores (see Table A.4). As a robustness check, we also study the correlation between management in the 200 control schools in our sample and endline test scores. Overall, the results show a higher correlation between management and test scores. A one-standard-deviation increase in the DWMS index is associated with an increase of 0.06σ–0.10σ in student test scores at endline among the control schools. Tables A.5 and A.6 provide more details. This higher correlation does not seem to be driven by sample differences, but rather by differences in the time of measurement, as the weak correlation between the DWMS and learning in 2015 is similar if we limit the sample to the 200 control schools (see Table A.7). Using the sample of control schools, Figure A.2 shows that there is high intertemporal correlation in learning outcomes and management practices, so a higher correlation between learning and management at endline is not driven by a systematic difference in how learning outcomes or management practices are measured between 2015 and 2018.

Table 4: Association between DWMS and test scores at baseline

	(1)	(2)	(3)	(4)	(5)
			PLANEA 2015 scores		
DWMS	0.0040	0.013	0.019	0.013	-0.0069
	(0.021)	(0.021)	(0.019)	(0.019)	(0.023)
No. of obs.	26,415	26,415	26,415	25,715	25,715
State FE	No	Yes	Yes	Yes	Yes
Strata FE	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
Enumerator FE	No	No	No	No	Yes

This table presents the conditional correlation between the DWMS and student test scores at baseline across all schools in our sample. State FE indicates whether state fixed effects are included. Strata FE indicates whether strata fixed effects are included. Controls indicates whether the regression controls for the number of pupils in the school, the pupil/teacher ratio, and the marginalization index. Enumerator FE indicates whether interviewer dummies are included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 Experimental results

Our main estimating equations for student-level outcomes are:

$$Y_{isg} = \alpha_g + \gamma_1 PECPlus_s + \varepsilon_{isg} \quad | \quad s \in \{PECplus, PEC\}, \tag{1}$$

$$Y_{isg} = \alpha_g + \gamma_2 PECPlus_s + \varepsilon_{isg} \quad | \quad s \in \{PECplus, Control\}, \tag{2}$$

$$Y_{isg} = \alpha_g + \gamma_3 PEC_s + \varepsilon_{isg} \quad | \quad s \in \{PEC, Control\}, \tag{3}$$

where Y_{isg} is the outcome of interest of student i in school s in group g (denoting the stratification group used to assign treatment), α_g are strata fixed effects; PEC_s is an indicator variable for a school s receiving grants, $PECPlus_s$ indicates a school s that received the management (plus grants) intervention, and ε_{isg} is an error term. We use a similar specification without i subscript to examine school-level outcomes. We estimate these models using ordinary least squares, clustering the standard errors at the school level.

Equation 1 is estimated using only PEC and PEC Plus schools. Thus, γ_1 reflects the difference between the "direct training" management intervention and a cascade-type intervention. Since all states randomly assigned schools to PEC and PEC Plus, all states inform this coefficient.

Equation 2 is estimated using only PEC Plus and control schools. Thus, γ_2 captures the effect of the "direct training" management intervention and the school grants compared to the control (which receives the management intervention through the cascade model). The identification of γ_2 is based on information from Puebla and Estado de México, the only two states that have control schools in their experiments.

Finally, Equation 3 is estimated using only PEC and control schools. Therefore, γ_3 captures the grants' effect compared to the control. Both PEC and control schools received the management intervention through the cascade model. This specification is estimated using only schools in Puebla, where we can compare PEC schools with control schools (the controls in Estado de México can only be compared to PEC Plus schools).

Table 5: Effects on the DWMS and its components

	(1) DWMS	(2) Operations	(3) Monitoring	(4) Targets	(5) People
Panel A : PEC Plus vs. PEC					
γ_1	0.12**	0.13**	0.13**	0.027	0.093*
	(0.052)	(0.055)	(0.060)	(0.053)	(0.056)
No. of obs.	913	913	913	913	913
Panel B : PEC Plus vs. Control					
γ_2	0.26***	0.26**	0.28**	0.15*	0.079
	(0.095)	(0.11)	(0.11)	(0.083)	(0.096)
No. of obs.	290	290	290	290	290
Panel C : PEC vs. Control					
γ_3	0.160	0.0753	0.186	0.0562	0.169
	(0.125)	(0.147)	(0.145)	(0.122)	(0.119)
No. of obs.	165	165	165	165	165
Panel D: Differences					
$\gamma_4 = \gamma_1 - \gamma_2$	-0.132	-0.127	-0.150	-0.124	0.015
p-value $(H_0 : \gamma_4 = 0)$	0.198	0.253	0.171	0.153	0.879
$\gamma_5 = \gamma_1 - \gamma_3$	-0.036	0.059	-0.057	-0.029	-0.075
p-value $(H_0 : \gamma_5 = 0)$	0.941	0.693	0.704	0.823	0.558
$\gamma_6 = \gamma_2 - \gamma_3$	0.096	0.186	0.093	0.095	-0.090
p-value $(H_0 : \gamma_6 = 0)$	0.356	0.171	0.490	0.436	0.462
p-value $(H_0 : \gamma_1 = \gamma_2 = \gamma_3)$	0.380	0.282	0.368	0.332	0.760

This table presents the treatment effects on management practices (measured using the DWMS). The outcome in Column 1 is the composite index of management practices, while Columns 2–5 display the outcomes for individual components of the management index. All regressions take into account the randomization design (i.e., they include strata fixed effects) and include enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Effects on learning outcomes

	(1) Math	(2) Language	(3) Average	(4) PCA
Panel A : PEC Plus vs. PEC				
γ_1	0.031	0.027	0.034	0.034
	(0.029)	(0.027)	(0.029)	(0.029)
No. of obs.	39,263	39,665	37,958	37,958
Panel B : PEC Plus vs. Control				
γ_2	0.051	0.036	0.047	0.047
	(0.048)	(0.041)	(0.046)	(0.046)
No. of obs.	15,802	15,930	15,474	15,474
Panel C : PEC vs. Control				
γ_3	-0.0790	-0.0745	-0.0832	-0.0829
	(0.0760)	(0.0685)	(0.0751)	(0.0749)
No. of obs.	6,799	6,798	6,673	6,673
Panel D: Differences				
$\gamma_4 = \gamma_1 - \gamma_2$	-0.020	-0.009	-0.012	-0.012
p-value $(H_0 : \gamma_4 = 0)$	0.687	0.838	0.803	0.806
$\gamma_5 = \gamma_1 - \gamma_3$	0.110	0.102	0.118	0.117
p-value $(H_0 : \gamma_5 = 0)$	0.202	0.200	0.171	0.172
$\gamma_6 = \gamma_2 - \gamma_3$	0.130	0.111	0.130	0.129
p-value $(H_0 : \gamma_6 = 0)$	0.071	0.100	0.073	0.074
p-value $(H_0 : \gamma_1 = \gamma_2 = \gamma_3)$	0.184	0.256	0.197	0.199

This table presents the treatment effects on learning outcomes (measured using PLANEA scores). The outcomes are math test scores (Column 1), language test scores (Column 2), the average across subjects (Column 3), and a composite index across subjects (Column 4). All regressions take into account the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2.1 The effect of the school grants (PEC versus controls)

We first examine the extent to which the grants affected management practices and learning outcomes (γ_3 above). As mentioned, Puebla is the only state in which schools were assigned to either grants (PEC) or control in the same experiment. Thus, the results in this section are entirely based on schools from this state.

The grants had no statistically significant impact on management practices (see Table 5, Panel C). The point estimate of the treatment effect is .16σ (p-value .2). We are able to rule out an effect greater than .41σ at the 95% level. This is unsurprising since there was no reason, a priori, that a monetary transfer conditional on a school improvement plan would improve managerial practices among school principals.

Next, we examine the effects of school grants in test scores two years after the program began (Table 6, Panel C). Students in PEC schools scored -0.08σ lower than those in control schools (p-value 0.27). We can rule out a positive effect greater than 0.06σ at the 95% level. This result is robust to a series of student- and school-level controls (see Table A.8). If anything, there is suggestive evidence that the treatment effect is negative and after including controls we can rule out an effect greater than 0.01σ at the 95% level.

Our results are consistent with previous findings from low-income countries that increasing schools' resources alone has no impact on student learning outcomes (e.g., Glewwe et al. (2009) in Kenya, Blimpo et al. (2015) in The Gambia, Das et al. (2013) in India, Pradhan et al. (2014) in Indonesia, Sabarwal et al. (2014) in Sierra Leone, and Mbiti et al. (2019) in Tanzania). Experimental evidence from the Mexican state of Colima shows that school grants had no impact on student learning outcomes one year after the intervention (Garcia-Moreno et al., 2019). We add to this literature and show a similar null effect of school grants in the Mexican state of Puebla two years after the start of implementation.

4.2.2 The effect of a high-quality management intervention (PEC Plus versus PEC)

As mentioned above, comparing PEC Plus to PEC (γ_1 above) isolates the effect of the high- quality intervention (implemented via a group of professional trainers) versus the same intervention implemented through the cascade model.

Overall, the direct management intervention improved management practices (see Table 5: Panel A). In PEC Plus schools management scores are .12σ (p-value 0.018) higher than in PEC schools. Therefore, our results show that investing in professional trainers to improve school principals' management capacities pays off.

Yet, while management practices improved as a result of the intervention, test scores did not (see Table 6: Panels A). Students in PEC Plus schools scored 0.03σ (p-value 0.24) higher than those in PEC schools. We can rule out, at the 95% level, that test scores increased by more than 0.09σ with respect to PEC schools. This result is robust to a series of student- and school-level controls (see Table A.8). After including controls we can rule out an effect greater than 0.08σ at the 95% level.

4.2.3 The combined effect of the high-quality management intervention and the school grants (PEC Plus versus the control group)

Comparing PEC Plus and control schools (γ_2 above) identifies the combined effect of receiving a high-quality management intervention and a school grant. These results are identified from schools in Estado de México and Puebla. In PEC Plus schools management scores are .26σ (p-value .0074) higher than in the control schools (see Table 5: Panel B). However, although test scores are higher in PEC Plus schools, the difference is not statistically significant (see Table 6: Panels A). Students in PEC Plus schools scored 0.05σ (p-value 0.32) higher than those in control schools. We can rule out, at the 95% level, that test scores increased by more than 0.14σ with respect to control schools. This result is robust to a series of student- and school-level controls (see Table A.8). After including controls we can rule out an effect greater than 0.08σ at the 95% level.

4.3 Complementarities between grants and management

The *Escuela al Centro* was designed to create synergies by providing schools with some (financial) autonomy and improving school principals' management practices.

Focusing on the management outcomes, the difference between PEC Plus schools and the control (γ_2) is larger than both the difference between PEC Plus and PEC (γ_1) and the difference between PEC and control schools (γ_3). However, neither difference $\gamma_1 - \gamma_2$ nor $\gamma_2 - \gamma_3$ is statistically significant (see Table 5: Panel D).

Turning to learning outcomes, the difference between PEC Plus schools and the control (γ_2) is larger than the difference between PEC Plus and PEC (γ_1) and the difference between PEC and control schools (γ_3). While $\gamma_1 - \gamma_2$ is not statistically significant, $\gamma_2 - \gamma_3$ is (p-values range from 0.073 to 0.1) (see Table 6: Panel D).

While we cannot formally test for complementarities between the programs, these results suggest synergies between high-quality managerial training and school grants: both inputs reinforce a positive effect on school principals' management practices and learning outcomes.[20]

4.4 Relationship between management and learning

As mentioned above, Bloom, Lemos, et al. (2015) find that a one-standard-deviation increase in the WMS index is associated with an increase in pupil outcomes of 0.2σ–

[20] To formally test for complementarities, we would need a treatment group that receives the "direct training" management intervention without the grants. See (Mbiti et al., 2019) for a detailed discussion of the experimental design needed to test complementarities.

0.4σ. The evidence from our baseline shows a weaker correlation between management practices and test scores.[21] Thus, being optimistic and assuming that the treatment effect on student learning of increasing management practices by one standard deviation is 0.4σ, an increase of .12σ in management practices (PEC Plus vs PEC) should yield an increase in test scores of .05σ (the actual treatment effect was 0.03σ) and an increase of .26σ in management practices (PEC Plus vs control) should yield an increase in test scores of .1σ (the actual treatment effect was 0.05σ).

Overall, the expected treatment effects on learning outcomes (given the treatment effects on management practices) are of the same order of magnitude as the actual treatment effects. While the intervention improved management practices, these improvements did not generate changes in learning outcomes that are statistically significant (even with a sample size of over 1,400 schools).[22] We cannot rule out the possibility that management did have a *small* positive impact on learning.

4.5 Relationship between management practices, the Stallings classroom observation and SisAT tools

Was it reasonable to expect that providing training on two specific tools, the Stallings classroom observation and SisAT, would improve managerial practices and test scores? We address this question by looking at the correlation between the self-reported information on the use of the Stalling classroom observation and SisAT tools on both the DWMS and test scores. These results, however, should be interpreted with caution since they rely on the self-reported assessments of school principals. Schools that answered the online surveys are statistically different from those that did not in almost every observable characteristic (see Tables A.13 and A.14).[23]

First, using the Stallings and SisAT tools is positively correlated with the DWMS (see Table A.9). In other words, the more likely principals are to use the management tools, the higher the DWMS index. Second, PEC Plus schools are more likely to implement the Stallings classroom observation tool and the SisAT tool (see Table A.12). Thus, the "direct training" management intervention is more successful than the cascade intervention at encouraging principals to actually use the management tools. Combining these two

[21] The correlation using the endline data among controls is similar to the one found by Bloom, Lemos, et al. (2015) in Brazil.

[22] Nor did *Escuela al Centro* have a statistically significant effect on other outcomes such as grade repetition and enrollment rates (see Table A.11).

[23] This section does not attempt to establish a causal relationship between the use of management tools, Stallings and SisAT, and the DWMS or test scores. The correlations included in this section are presented for completeness.

results—PEC Plus schools are more likely to use the management tools provided to them, and these tools are correlated with the DWMS—it is unsurprising that the PEC Plus improves DWMS scores (as shown in Table 5). Finally, the self-reported information also shows that the correlation between using the management tools and test scores is not statistically significant (see Table A.10), which is aligned with the finding that PEC Plus schools did not improve their learning outcomes (as shown in Table 6).

4.6 Heterogeneity

This section explores heterogeneous treatment effects on management practices by schools' (and school principals') baseline characteristics. Overall, there is little evidence of heterogeneity.

Specifically, we estimate the following equations:

$$Y_{isg} = \alpha_g + \beta_1 PECPlus_s + \beta_2 PECPlus_s \times c_s + \beta_3 c_s + \varepsilon_{isg} \quad | \quad s \in \{PECplus, PEC\}, \tag{4}$$

$$Y_{isg} = \alpha_g + \beta_1 PECPlus_s + \beta_2 PECPlus_s \times c_s + \beta_3 c_s + \varepsilon_{isg} \quad | \quad s \in \{PECplus, Control\}, \tag{5}$$

$$Y_{isg} = \alpha_g + \beta_1 PEC_s + \beta_2 PEC_s \times c_s + \beta_3 c_s + \varepsilon_{isg} \quad | \quad s \in \{PEC, Control\}, \tag{6}$$

where c_s denotes the school characteristics along which we wish to measure heterogeneity, and β_2 allows us to test whether there is any differential treatment effect. Everything else is as in Equations 1-3. We study heterogeneity by the school's baseline management quality, the school's marginalization index, and the school principal's gender and tenure.

Focusing on the DWMS, we find no heterogeneity in treatment effects (see Table A.15). Focusing on learning outcomes, there is some evidence that school grants have a negative treatment effect when the principal is male, and a null effect when the principal is female (see Table A.16). However, since we did not pre-specify these hypotheses, we report the results for completeness and leave it to future work to explicitly test whether the school principal's gender is an important factor in school grants' effectiveness.

5 Cost-effectiveness analysis

In this section, we estimate the cost of implementing the interventions and assess their cost-effectiveness vis-à-vis the gains in management quality and learning outcomes. From an education policy perspective, this analysis is useful for understanding the costs and benefits of scaling up each intervention (school grants, managerial practices or both).

Our analysis includes both variable and fixed costs. The data we use to estimate the costs come from four sources: (i) official program documents; (ii) interviews with implementing staff from the federal government; (iii) administrative records; and (iv) federal government payroll data.[24] We do not include the direct and indirect costs of implementing the overall *Escuela al Centro* strategy, beyond those associated with the PEC and PEC Plus interventions. All cost estimates reported are expressed in 2015 USD and correspond to average per-school costs.

Table 7 summarizes the interventions' total costs, dividing them into variable and fixed costs (staff time). The variable costs include: (i) school grants and (ii) managerial training. We report the average variable cost of implementing the interventions across the seven participating states. The total fixed cost of implementing the interventions in PEC and PEC schools (1,466 schools) over the two years of implementation was approximately 2,143,779 USD (1,462 USD per school). This cost includes the salaries of the federal- and state-level field coordination teams, which led the design, implementation, and monitoring of all activities included as part of this impact evaluation. Section A.1, in the appendix, provides more details on the cost calculations.

[24] All federal employees' salaries in Mexico are published annually on the *Nomina Transparente* website: https://nominatransparente.rhnet.gob.mx/.

Table 7: Total cost of implementing the interventions over two years by treatment group (in 2015 USD)

Description	Per-school cost			Per-student cost		
	PEC Plus	PEC	Control	PEC Plus	PEC	Control
Panel A: Fixed cost[a]						
Staff time	1,462	1,462	0.0	5.23	5.23	0.00
Panel B: Variable costs[b]						
Managerial intervention	472	2.4	2.4	1.61	0.01	0.01
School grant	8,718	8,484	0.0	29.7	32.1	0.00
Total	10,652	9,948	2.4	36.6	37.4	0.13

[a] The core team from the Federal Education Authority included: one Director General, one Coordinator of Academic Activities, one Head of Technical Issues and Monitoring, and two Department Chiefs. The core team in each state-level education authority generally included three staff members at the "Subdirector de Area" level. The core team that developed the classroom observation tool included five senior staff from the Federal Education Authority—one "Director de Area," one "Subdirector de Area," two "Director de departamento," and one analyst. The tool was adapted from an existing classroom observation tool that was developed in 2011 as part of a large-scale study on classroom time use (Bruns & Luque, 2014). The core team for developing the data to guide the instruction tool included eight senior staff from the Federal Education Authority—one "Director de Area," five "Subdirector de Area," and two IT experts. The tool was developed for the implementation of the *Escuela al Centro* strategy and it involved close coordination with staff from the Curriculum Area of the Federal Education Authority.

[b] The per-school variable cost in PEC and control schools is the same because they both benefited from the cascade training.

Since none of the interventions studied in this paper improved PLANEA scores, they are not cost-effective for improving learning outcomes. However, they may still be of interest if the objective is to improve managerial practices.

First, we compare PEC schools to control schools (i.e., the effect of the grant). Since PEC schools exhibit no improvement in management practices, the grants alone are not cost-effective. The impact of the managerial intervention delivered by professional trainers (PEC Plus) relative to PEC on the DWMS is 0.12 standard deviations over two years at a total cost (relative to PEC) of 470 USD per school (1.60 USD per student). Overall, having a managerial training strategy delivered by professional trainers is cost-effective to improve management practices. The PEC Plus treatment (managerial and grant interventions) increased the DWMS by 0.26 standard deviations after two years at a total cost of 10,652 USD per school (36.60 USD per student). Thus, even though there is some evidence of synergies between the high-quality management intervention and the school grants, the grants' cost is so large that this calls into question whether the additional gains in terms of managerial practices are justified.

Overall, our results suggest that investing in professional trainers to improve school principals' management capacities can be cost-effective if there is already an investment in place to provide schools with financial autonomy (i.e., school grants). However, expanding both financial autonomy (providing grants) and improving managerial practices is not cost-effective given the large cost of the grants, and the fact that grants on their own have no effect on test scores or management practices.

6 Conclusions

Recent studies have identified the pivotal role that managerial practices play in helping an organization reach its objectives (Bender, Bloom, Card, Van Reenen, & Wolter, 2018), and the education sector is no exception. This paper reports some of the first experimental evidence of the effects of an intervention to improve school management in a developing country. We randomly assigned a group of public primary schools in seven states in Mexico to one of three treatments: receiving a school grant (PEC), receiving a grant plus training to monitor basic skills and perform classroom observations (PEC Plus), and a control group. Our results show that the grant, by itself, had no impact on either school principals' managerial capacities (measured by the DWMS) or student learning (measured by the PLANEA). A well-implemented training strategy improved school principals' managerial capacity, and our results identified some synergies between the training strategy and the school grants. PEC Plus schools improved their management capacity by $.12\sigma$ (p-value 0.018) compared to PEC and $.26\sigma$ (p-value .0074) compared to control schools. Although positive, the effects of the interventions on student learning are not statistically different from zero.

Our results suggest that while it is possible to improve managerial practices among school principals, this requires a well-implemented strategy, with professional trainers instead of a cascade model. To improve student learning in the short term, a management intervention may need a larger impact on school principals' managerial capacities.

References

Álvarez, J., García-Moreno, V., & Patrinos, H. A. (2007). *Institutional effects as determinants of learning outcomes: Exploring state variations in mexico* (Vol. 4286). World Bank Publications.

Bender, S., Bloom, N., Card, D., Van Reenen, J., & Wolter, S. (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics*, *36*(S1), S371-S409. Retrieved from https://doi.org/10.1086/694107 doi: 10.1086/694107

Blimpo, M. P., Evans, D. K., & Lahire, N. (2015). *Parental human capital and effective school management : Evidence from The Gambia* (Policy Research Working Paper No. 7238). World Bank.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics*, *128*(1), 1–51.

Bloom, N., Lemos, R., Sadun, R., Scur, D., & Van Reenen, J. (2014). The new empirical economics of management. *Journal of the European Economic Association*, *12*(4), 835–876.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic Journal*, *125*(584), 647–674.

Bloom, N., Propper, C., Seiler, S., & Van Reenen, J. (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *Review of Economic Studies*, *82*(2), 457-489.

Bloom, N., & Van Reenen, J. (2007, 11). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, *122*(4), 1351-1408. doi: 10.1162/qjec.2007.122.4.1351

Brunetti, A., Büchel, K., Jakob, M., Jann, B., Kühnhanss, C., & Steffen, D. (2020, March). Teacher Content Knowledge in Developing Countries: Evidence from a Math Assessment in El Salvador [University of Bern Social Sciences Working Papers]. (34). Retrieved from https://ideas.repec.org/p/bss/wpaper/34.html

Bruns, B., & Luque, J. (2014). *Great teachers: How to raise student learning in Latin America and the Caribbean*. World Bank Publications.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, *5*(2), 29–57.

de Hoyos, R., Ganimian, A. J., & Holland, P. A. (2019, 11). Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity Building for Public Schools in Argentina. *The World Bank Economic Review*.

de Hoyos, R., Ganimian, A. J., & Holland, P. A. (2020). Great things come to those who wait: Experimental evidence on performance-management tools and training in public schools in argentina.

de Hoyos, R., García-Moreno, V., & Patrinos, H. A. (2017). The impact of an accountability intervention with diagnostic feedback: Evidence from mexico. *Economics of*

Education Review, *58*, 123 - 140.

Dirección General de Planeación, Programación y Estadística Educativa. (2018). *Sistema Educativo de los Estados Unidos Mexicanos, Principales Cifras 2017-2018* (Tech. Rep.). Secretaría de Educación Pública. Retrieved from https://www.planeacion.sep.gob.mx/estadisticaeindicadores.aspx

Dobbie, W., & Fryer, R. (2013, October). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, *5*(4), 28-60. doi: 10.1257/app.5.4.28

Elacqua, G., Iribarren, M. L., & Santos, H. (2018). *Private schooling in latin america: Trends and public policies* (Tech. Rep.). Inter-American Development Bank. doi: http://dx.doi.org/10.18235/0001394

Fryer, R. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics*, *129*(3), 1355–1407.

Fryer, R. (2017, May). *Management and student achievement: Evidence from a randomized field experiment* (Working Paper No. 23437). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w23437 doi: 10.3386/w23437

Garcia-Moreno, V., Gertler, P., & Patrinos, H. A. (2019). *School-based management and learning outcomes* (Tech. Rep.). World Bank, Washington, DC.

Gilligan, D. O., Karachiwalla, N., Kasirye, I., Lucas, A. M., & Neal, D. (2019). Educator incentives and educational triage in rural primary schools. *Journal of Human Resources*, 1118–9871R2.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, *1*(1), 112–35.

INEGI. (2018). *Marco geoestadístico de Mexico.* Retrieved 06/01/2018, from https://www.inegi.org.mx/temas/mg/default.html#

Instituto Nacional para la Evaluación de la Educación. (2018). *Resultados de PLANEA* (Tech. Rep.). Retrieved from https://www.inee.edu.mx/evaluaciones/planea/resultados-planea/

Lemos, R., Muralidharan, K., & Scur, D. (2021, January). *Personnel management and school productivity: Evidence from india* (Working Paper No. 28336). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w28336 doi: 10.3386/w28336

Lemos, R., & Scur, D. (2016). Developing management: An expanded evaluation tool for developing countries. *RISE Working Paper, 16(007)*.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019, 04).

Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *The Quarterly Journal of Economics*, *134*(3), 1627-1673. Retrieved from https://doi.org/10.1093/qje/qjz010 doi: 10.1093/qje/qjz010

Muralidharan, K., & Singh, A. (2020, November). *Improving public sector management at scale? experimental evidence on school governance india* (Working Paper No. 28129). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w28129 doi: 10.3386/w28129

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2018). *Teacher professional development around the world: The gap between evidence and practice* (Tech. Rep.). The World Bank.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., & Artha, R. P. (2014, April). Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia. *American Economic Journal: Applied Economics*, *6*(2), 105-26.

Sabarwal, S., Evans, D. K., & Marshak, A. (2014). *The permanent input hypothesis : the case of textbooks and (no) student learning in Sierra Leone* (Policy Research Working Paper Series No. 7021). The World Bank.

Secretaría de Educación Pública, & Banco Internacional de Reconstrucción y Fomento. (2015). *Evaluación de Impacto del Ejercicio y Desarrollo de la Autonomía de Gestión Escolar y Estrategia de Intervención Controlada* (Tech. Rep.). Retrieved from https://escuelaalcentro.com/wp-content/uploads/2018/04/Documento-base-Evaluaci%25c3%25b3n-de-Impacto.pdf

Stallings, J. (1977). *Learning to look: a handbook on classroom observation and teaching models*. Wadsworth Pub. Co. Retrieved from https://books.google.com.mx/books?id=QEglAQAAIAAJ

Stallings, J., & Molhlman, G. (1988). Classroom observation techniques. In J. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook.* Elsevier Science & Technology Books.

World Bank. (2007). *What is school-based management?*

World Bank. (2017a). *Primary completion rate, total (% of relevant age group).* (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.PRM.CMPT.ZS?locations=MX)

World Bank. (2017b). *School enrollment, primary (% net).* (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.PRM.NENR?locations=MX)

A Online Appendix for "School management, grants, and test scores: Experimental Evidence from Mexico" by Bedoya, de Hoyos, Romero, Silveyra and Yanez-Pagans

Table A.1: Balance statistics across evaluation participants from other schools

	(1)	(2)	(3)
	Mean		Difference
Variable	Participant	Non-participant	(1)-(2)
Students in math achievement L-IV (%)	7.71	9.76	-2.05***
	(11.16)	(17.41)	(0.28)
Students in math achievement L-I (%)	60.18	56.70	3.47***
	(21.32)	(28.34)	(0.53)
Students in language achievement L-IV (%)	3.21	5.63	-2.42***
	(5.05)	(11.26)	(0.13)
Students in language achievement L-I (%)	48.91	43.44	5.47***
	(20.37)	(28.27)	(0.51)
Marginalization	0.57	0.38	0.20***
	(0.49)	(0.48)	(0.01)
Urbanization	0.41	0.34	0.07***
	(0.49)	(0.47)	(0.01)
Number of students	302.82	165.64	137.18***
	(182.26)	(175.11)	(4.51)
Number of teachers	10.32	6.31	4.01***
	(5.07)	(6.22)	(0.13)
Student-teacher ratio	28.58	26.60	1.98***
	(7.12)	(11.49)	(0.18)
Observations	1,650	120,519	122,169

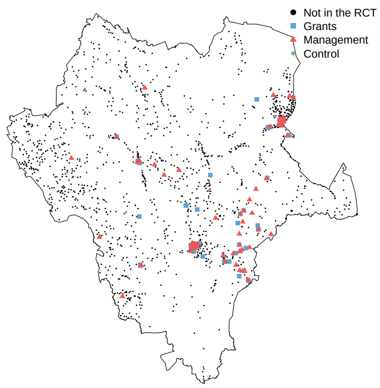
This table presents the mean and standard error of the mean (in parentheses) for schools not in the experiment (Column 1) and schools in the experiment (Column 2). Column 3 shows the mean difference between participant and non-participant schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to one of four knowledge domains granted to students in the results of the PLANEA exam. L-I refers to the lowest level, while L-IV refers to the highest level. Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is coded 1 for schools located in an urban area, and 0 otherwise. The number of students and teachers is taken from Formato 911 from the year 2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Balance statistics across backscored & Non-backscored interviews

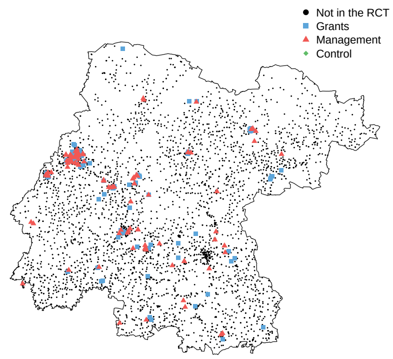
	(1)	(2)	(3)
	Mean		**Difference**
Variable	Non-DWMS Audio	DWMS Audio	(1)-(2)
Students in math achievement L-IV (%)	8.07	8.89	-0.82
	(11.61)	(11.00)	(0.68)
Students in math achievement L-I (%)	59.59	57.95	1.64
	(21.35)	(22.47)	(1.35)
Students in language achievement L-IV (%)	3.07	3.41	-0.34
	(5.14)	(5.28)	(0.32)
Students in language achievement L-I (%)	50.89	48.95	1.95
	(20.17)	(21.05)	(1.27)
Marginalization	0.58	0.53	0.05
	(0.49)	(0.50)	(0.03)
Urbanization	0.41	0.48	-0.07**
	(0.49)	(0.50)	(0.03)
Number of students	294.71	309.80	-15.09
	(176.98)	(196.80)	(11.74)
Number of teachers	9.77	10.27	-0.50*
	(4.62)	(4.93)	(0.30)
Student-teacher ratio	29.18	28.73	0.45
	(6.95)	(7.28)	(0.44)
Observations	350	1,142	1,492

This table presents the mean and standard error of the mean (in parentheses) for schools without audio for the DWMS endline interview (Column 1) and schools with it (Column 2). Column 3 shows the mean difference between backscored and non backscored schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to one of four knowledge domains granted to students in the results of the PLANEA exam. L-I refers to the lowest level, while L-IV refers to the highest level. Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from Formato 911 for the year 2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

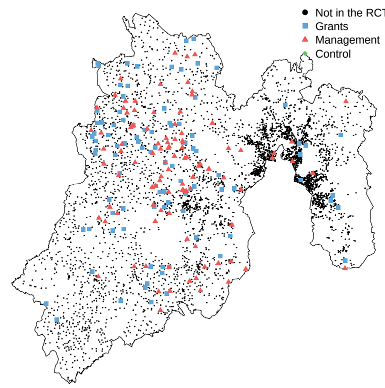
Figure A.1: Geographical distribution of the treatment assignment by state



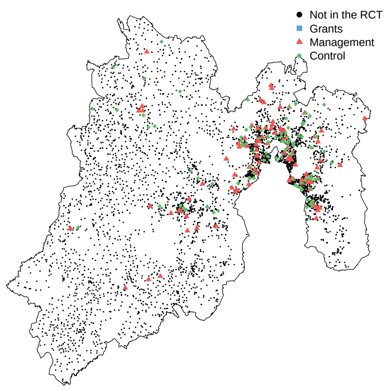
(a) Durango



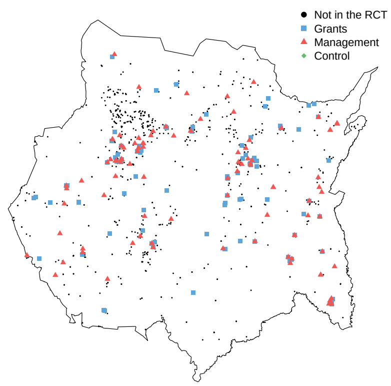
(b) Guanajuato



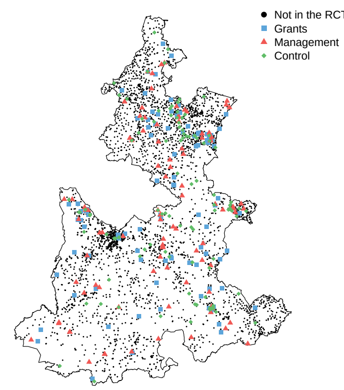
(c) Estado de México – 1



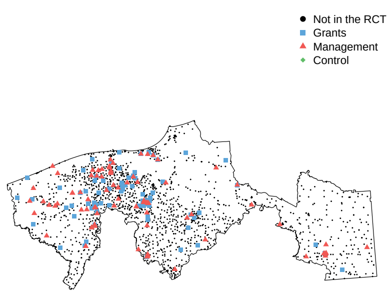
(d) Estado de México – 2



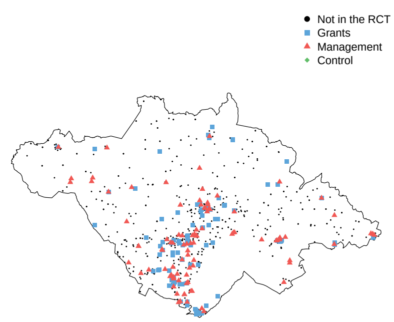
(e) Morelos



(f) Puebla



(g) Tabasco



(h) Tlaxcala

Table A.3: Treatment assignment by state

State	PEC	PEC Plus	Control	Blocking
Durango	100	100	0	None.
Estado de México –1	100	100	0	None.
Estado de México –2	0	100	100	None.
Morelos	100	100	0	Blocking by school level. 130 primary schools and 70 secondary schools in separate blocks.
Tlaxcala	100	100	0	Blocking by school level. 110 primary schools and 90 non-primary schools in separate blocks.
Guanajuato	100	100	0	200 schools randomly sampled. Schools were ranked based on enrollment. Odd-ranked schools were assigned to PEC and even-ranked schools were assigned to PEC Plus.
Tabasco	83	82	0	165 eligible schools (with 60 or more students and 5 or more teachers). Eligible schools were ranked by priority (level 2, 4, 5 and 6) and number of students. Odd ranked schools were assigned to PEC and even ranked schools were assigned to PEC Plus.
Puebla	101	100	99	300 eligible schools (60 or more students and 5 or more teachers). Blocks were created based on marginalization and urbanization. For the marginalization categories, the “very high” and “high” CONAPO classifications were grouped into a single category (“high” marginalization) and the “medium,” “low” and “very low” CONAPO classifications were grouped into another single category (“not high” marginalization). Thus, schools were assigned to four groups based on marginalization (high/low) and whether they were urban or rural. Schools were ranked within each group based on enrollment. Schools were assigned in a cycle to PEC, PEC Plus, or control. The order in which the cycle began was randomized.

This table shows the number of schools and, for each design, their distribution among treatment conditions. Note there are two distinct designs for Estado de México, both of which have 200 schools. Blocking was used in five of the eight designs. While in some states the randomization included secondary schools, in this paper we focused on primary schools as this is the sample for which we have learning outcome data. Since the randomization, in such cases, was stratified by school type (primary vs. secondary), this sample restriction does not invalidate the internal validity of our results.

Table A.4: Association between DWMS components and test scores at baseline across all schools

	(1)	(2)	(3)	(4)
	PLANEA 2015 scores			
Operations	-0.014			
	(0.022)			
Monitoring		0.000075		
		(0.021)		
Targets			0.0074	
			(0.022)	
People				-0.016
				(0.025)
No. of obs.	25,715	25,715	25,715	25,715

This table presents the conditional correlation between DWMS components and student test scores at baseline across all schools. All regression controls for strata fixed effects, the number of pupils in the school, the pupil/teacher ratio, the marginalization index, and enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Association between DWMS and test scores in control school at endline

	(1)	(2)	(3)	(4)	(5)
	PLANEA 2018 scores: control schools				
DWMS	0.056	0.058	0.063*	0.059*	0.096
	(0.036)	(0.036)	(0.035)	(0.032)	(0.059)
No. of obs.	5,521	5,521	5,521	5,454	5,454
State FE	No	Yes	Yes	Yes	Yes
Strata FE	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
Enumerator FE	No	No	No	No	Yes

This table presents the conditional correlation between DWMS and student test scores in 2018 across the 200 control schools in our sample (100 from Estado de México and 100 from Puebla). State FE indicates whether state fixed effects are included. Strata FE indicates whether strata fixed effects are included. Controls indicates whether the regression controls for the number of pupils in the school, the pupil/teacher ratio, and the marginalization index. Enumerator FE indicates whether interviewer dummies are included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Association between DWMS components and test scores in control schools at endline

	(1)	(2)	(3)	(4))
	PLANEA 2018 scores			
Operations	-0.013			
	(0.047)			
Monitoring		0.100^{**}		
		(0.046)		
Targets			0.13^{**}	
			(0.062)	
People				0.039
				(0.056)
No. of obs.	5,454	5,454	5,454	5,454

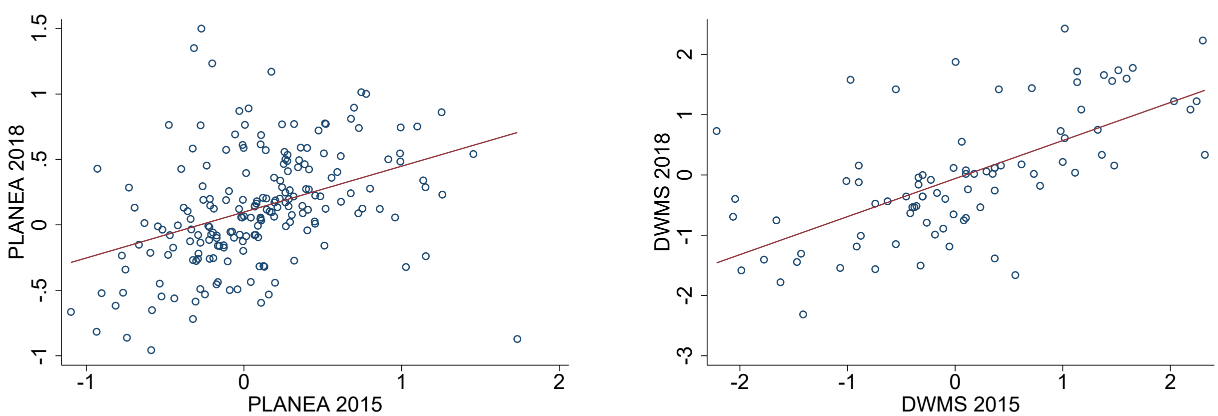
This table presents the conditional correlation between DWMS components and student test scores in 2018 across the 200 control schools in our sample (100 from Estado de México and 100 from Puebla). All regression controls for strata fixed effects, the number of pupils in the school, and the pupil/teacher ratio, the marginalization index, and enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Association between DWMS and test scores in control school at baseline in control schools

	(1)	(2)	(3)	(4)	(5)
	PLANEA 2015 scores: control schools				
DWMS	0.0032	0.0034	0.0087	-0.0023	0.036
	(0.049)	(0.049)	(0.045)	(0.041)	(0.047)
No. of obs.	3,702	3,702	3,702	3,633	3,633
State FE	No	Yes	Yes	Yes	Yes
Strata FE	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
Enumerator FE	No	No	No	No	Yes

This table presents the conditional correlation between DWMS and student test scores in 2015 across the 200 control schools in our sample (100 from Estado de México and 100 from Puebla). State FE indicates whether state fixed effects are included. Strata FE indicates whether strata fixed effects are included. Controls indicates whether the regression controls for the number of pupils in the school, the pupil/teacher ratio, and the marginalization index. Enumerator FE indicates whether interviewer dummies are included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.2: Intertemporal correlation of learning outcomes and management practices in control school



(a) Learning outcomes

(b) Management practices

Table A.8: Effects on learning outcomes

	(1)	(2)	(3)	(4)
	PCA score			
Panel A : PEC Plus vs. PEC				
γ_1	0.034	0.032	0.032	0.035
	(0.029)	(0.025)	(0.025)	(0.024)
No. of obs.	37,958	37,958	37,958	37,958
Panel B : PEC Plus vs. Control				
γ_2	0.047	0.012	0.0065	0.0082
	(0.046)	(0.039)	(0.040)	(0.039)
No. of obs.	15,474	15,474	15,474	15,474
Panel C : PEC vs. Control				
γ_3	-0.0829	-0.104	-0.106*	-0.105*
	(0.0749)	(0.0630)	(0.0628)	(0.0590)
No. of obs.	6,673	6,673	6,673	6,673
Lagged scores	No	Yes	Yes	Yes
Student controls	No	No	Yes	Yes
School controls	No	No	No	Yes

This table presents the treatment effects on learning outcomes (measured using PLANEA scores). The outcome is a composite index across subjects. All regressions take into account the randomization design (i.e., include strata fixed effects). Lagged scores indicates whether school average test-scores from 2015 are included as controls. Student controls indicates whether age and gender are included as controls. School controls indicate whether the following controls are included: whether the school has a day shift, whether is a primary school intended to serve indigenous population, the school's age, whether the school is located in an urban area, and the marginalization index of the school's municipality. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Association between DWMS and implementation and use indexes

	(1)	(2)	(3)	(4)
	Management index			
Implementation index Stallings	0.035			
	(0.028)			
Use index Stallings		0.068***		
		(0.026)		
Implementation index SisAT			0.052**	
			(0.024)	
Use index SisAT				0.065**
				(0.029)
No. of obs.	765	758	879	872

This table presents the conditional correlation between DWMS and implementation and use indexes. Both the implementation and use indexes are constructed as the simple average of the online survey variables for each element of the intervention. The management index and the implementation and use indexes are standardized. All regression controls for strata fixed effects and enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Association between learning outcomes and implementation and use indexes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Math				**Language**				**Average**			
Implementation index Stallings	0.013				0.0020				0.0092			
	(0.013)				(0.013)				(0.014)			
Use index Stallings		0.0090				0.021				0.016		
		(0.015)				(0.015)				(0.016)		
Implementation index SisAT			0.0056				0.015				0.011	
			(0.013)				(0.012)				(0.013)	
Use index SisAT				-0.012				-0.021				-0.018
				(0.014)				(0.014)				(0.014)
No. of obs.	35,365	35,134	39,944	39,639	35,778	35,544	40,313	40,009	34,230	34,006	38,794	38,503

This table presents the conditional correlation between learning outcomes and implementation and use indexes. Both the implementation and use indexes are constructed as the simple average of the online survey variables for each element of the intervention. The learning outcomes and the implementation and use indexes are standardized. All regression controls for strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Effects on other outcomes

	(1) Pass rate	(2) Repetition rate	(3) Enrollment
Panel A : PEC Plus vs. PEC			
γ_1	0.0308	0.0251	13.65
	(0.253)	(0.114)	(8.734)
No. of obs.	1,186	1,185	1,192
Control mean	99.22	0.72	258.49
Panel B : PEC Plus vs. Control			
γ_2	0.662	-0.154	2.010
	(0.519)	(0.143)	(19.73)
No. of obs.	394	395	396
Control mean	98.89	0.71	326.44
Panel C : PEC vs. Control			
γ_3	0.915	0.0684	13.54
	(1.019)	(0.227)	(22.19)
No. of obs.	199	199	200
Control mean	98.10	0.87	259.96

This table presents the treatment effects on the percentage of students who approve their grade and can progress to the next one (pass rate in Column 1), the percentage of students that repeat a grade (Column 2), and the total number of students enrolled (Column 3). All outcomes refer to the 2017–2018 school cycle. All regressions take into account the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Treatment on Stallings and SisAT implementation and use

	Implementation index Stallings	Use index Stallings	Implementation index SisAT	Use index SisAT
Panel A : PEC Plus vs. PEC				
γ_1	0.38***	0.36***	0.15**	-0.080
	(0.072)	(0.070)	(0.067)	(0.066)
No. of obs.	827	822	866	860
Control mean	-0.26	-0.21	-0.08	0.08
Panel B : PEC Plus vs. control				
γ_2	0.68***	0.73***	0.41***	-0.057
	(0.17)	(0.13)	(0.10)	(0.11)
No. of obs.	258	255	352	350
Control mean	-0.37	-0.44	-0.17	-0.04
Panel C : PEC vs. Control				
γ_3	0.62	0.44	0.11	-0.031
	(0.54)	(0.33)	(0.15)	(0.14)
No. of obs.	15	13	153	149
Control mean	-0.64	-0.74	-0.13	0.21

Both the implementation and use indexes are constructed as the simple average of the online survey variables for each element of the intervention. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Balance statistics across evaluation schools taking the implementation and use Stallings' survey

	(1)	(2)	(3)
	Mean		**Difference**
Variable	Participant	Non-participant	(1)-(2)
Students in math achievement L-IV (%)	8.40	7.95	0.45
	(11.60)	(11.19)	(0.63)
Students in math achievement L-I (%)	59.33	58.93	0.41
	(21.38)	(22.15)	(1.22)
Students in language achievement L-IV (%)	3.26	2.92	0.34
	(5.46)	(4.49)	(0.27)
Students in language achievement L-I (%)	49.63	52.19	-2.57**
	(19.79)	(21.53)	(1.17)
Marginalization	0.55	0.62	-0.07**
	(0.50)	(0.49)	(0.03)
Urbanization	0.44	0.38	0.06**
	(0.50)	(0.49)	(0.03)
Number of students	306.01	281.43	24.59**
	(185.73)	(172.17)	(9.83)
Number of teachers	10.15	9.32	0.84***
	(4.80)	(4.41)	(0.25)
Student-teacher ratio	29.07	29.10	-0.03
	(7.12)	(6.82)	(0.39)
Observations	1,021	471	1,492

This table presents the mean and standard error of the mean (in parentheses) for schools taking the Stallings implementation and use survey (Column 1) and schools not taking it (Column 2). Column 3 shows the mean difference between participant and non-participant schools, as well as the standard error of the difference, clustered at the school level. Achievement Level (L) refers to one out of four knowledge domains granted to students in the results of the exam PLANEA. L-I refers to the lowest level, while L-IV refers to the highest level. Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from Formato 911 for the year 2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.14: Balance statistics across evaluation schools taking the implementation and use SisAT' survey

	(1)	(2)	(3)
	Mean		**Difference**
Variable	Participant	Non-participant	(1)-(2)
Students in math achievement L-IV (%)	8.43	7.66	0.77
	(11.37)	(11.80)	(0.73)
Students in math achievement L-I (%)	58.54	61.54	-3.00**
	(21.46)	(22.05)	(1.36)
Students in language achievement L-IV (%)	3.23	2.88	0.34
	(5.42)	(4.20)	(0.28)
Students in language achievement L-I (%)	49.95	52.15	-2.20*
	(20.35)	(20.44)	(1.27)
Marginalization	0.60	0.47	0.13***
	(0.49)	(0.50)	(0.03)
Urbanization	0.41	0.46	-0.05
	(0.49)	(0.50)	(0.03)
Number of students	304.44	276.55	27.89***
	(187.15)	(160.37)	(10.38)
Number of teachers	10.06	9.28	0.78***
	(4.82)	(4.18)	(0.27)
Student-teacher ratio	29.13	28.91	0.22
	(7.08)	(6.84)	(0.43)
Observations	1,161	331	1,492

This table presents the mean and standard error of the mean (in parentheses) for schools taking the SisAT implementation and use survey (Column 1) and schools not taking it (Column 2). Column 3 shows the mean difference between participant and non-participant schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to one of four knowledge domains granted to students in the results of the PLANEA exam. L-I refers to the lowest level, while L-IV refers to the highest level. Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from Formato 911 for the year 2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.15: Heterogeneous effects on management

	Management 2015	Principal's gender	Principal's tenure	Marginalization
Panel A : PEC Plus vs. PEC				
β_1	0.091	0.090	0.18***	0.21***
	(0.071)	(0.069)	(0.064)	(0.074)
β_2	0.070	0.080	-0.012	-0.16
	(0.070)	(0.10)	(0.0086)	(0.10)
β_3	0.070	0.080	-0.012	-0.16
	(0.070)	(0.10)	(0.0086)	(0.10)
No. of obs.	511	913	913	913
Control mean	0.12	-0.04	-0.04	-0.04
Panel B : PEC Plus vs. Control				
β_1	0.36***	0.25*	0.31**	0.37**
	(0.11)	(0.14)	(0.14)	(0.14)
β_2	0.084	0.0093	-0.0092	-0.16
	(0.13)	(0.20)	(0.018)	(0.21)
β_3	0.21*	0.17	0.010	0.34**
	(0.12)	(0.13)	(0.012)	(0.14)
No. of obs.	171	290	290	290
Control mean	-0.01	-0.05	-0.05	-0.05
Panel C : PEC vs. Control				
β_1	0.26	0.17	0.085	0.19
	(0.16)	(0.16)	(0.17)	(0.29)
β_2	-0.055	-0.070	0.021	-0.044
	(0.17)	(0.25)	(0.023)	(0.32)
β_3	0.21	0.20	0.012	-0.038
	(0.14)	(0.17)	(0.017)	(0.26)
No. of obs.	93	165	165	165
Control mean	-0.20	-0.19	-0.19	-0.19

This table shows the results from estimating Equations 4–6, when the outcome variable is the 2018 DWMS index. Management 2015 refers to the index calculated with baseline information, Principal's gender takes a value of 1 for female principals and 0 for males, Principal's tenure refers to the number of years as principal, and Marginalization takes a value of 1 for schools located in areas with high or very high marginalization. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.16: Heterogeneous effects on learning

	Management 2015	Principal's gender	Principal's tenure	Marginalization
Panel A : PEC Plus vs. PEC				
β_1	0.048	0.0076	0.010	0.047
	(0.042)	(0.037)	(0.037)	(0.044)
β_2	-0.052	0.056	0.0043	-0.027
	(0.042)	(0.059)	(0.0051)	(0.057)
β_3	0.036	0.070*	-0.00022	-0.21***
	(0.026)	(0.041)	(0.0034)	(0.043)
No. of obs.	19,112	37,958	37,867	37,958
Control mean	-0.05	-0.07	-0.07	-0.07
Panel B : PEC Plus vs. Control				
β_1	0.11	0.0018	0.035	0.068
	(0.069)	(0.066)	(0.065)	(0.064)
β_2	-0.12*	0.10	0.00036	-0.043
	(0.069)	(0.094)	(0.0076)	(0.090)
β_3	0.048	0.055	0.0059	-0.16***
	(0.042)	(0.058)	(0.0050)	(0.062)
No. of obs.	7,300	15,474	15,407	15,474
Control mean	0.13	0.15	0.15	0.15
Panel C : PEC vs. Control				
β_1	-0.11	-0.21**	-0.061	-0.081
	(0.098)	(0.091)	(0.093)	(0.19)
β_2	-0.015	0.29*	-0.0043	-0.0041
	(0.091)	(0.15)	(0.012)	(0.21)
β_3	0.080	-0.099	0.013	-0.31*
	(0.056)	(0.11)	(0.0083)	(0.17)
No. of obs.	3,860	6,673	6,649	6,673
Control mean	0.19	0.15	0.15	0.15

This table shows the results from estimating Equations 4–6, when the outcome variable is the PCA index from math and language 2018 PLANEA scores. Management 2015 refers to the index calculated with baseline information, Principal's gender takes a value of 1 for female principals and 0 for males, Principal's tenure refers to the number of years as principal, and Marginalization takes a value of 1 for schools located in areas with high or very high marginalization. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.1 Detailed cost calculations

A.1.1 School grants

The main variable costs were, by far, the school grants. Each PEC Plus school received on average 11,064 USD in school grants over two years (paid on an annual basis). The criteria for allocating grants changed during the two years of implementation. The criteria to allocate school grants is defined in the *Escuela al Centro* operational rules, which are revised on an annual basis. During the first year, the allocation of grants was based on state-level formulas, which included both equity and quality components. The equity component benefited all schools and was determined based on two indicators: (i) number of students in the school, with larger schools receiving a larger share of the resources and (ii) school locality marginalization level, with schools located in more marginalized areas receiving a larger share of the resources. The quality component benefited only schools that improved their educational indicators in the previous academic year. The quality indicators varied across states but generally included a combination of: (i) student standardized learning assessments; (ii) approval and retention rates; and (iii) terminal efficiency. We collected data ex-post from school principals participating in the impact evaluation to understand their perceptions of the use of formulas to allocate school grants. On average, 86% of principals in the seven participating states reported knowing about the use of formulas. However, 61% of those who knew about the formulas mentioned they could not understand how the formula worked and/or what indicators were monitored by their state-level education authority to allocate school grants.

Starting in the second year of implementation, the use of formulas to distribute school grants was discontinued and resources were distributed according to the number of students in each school.

A.1.2 School managerial training

The second-largest variable cost was the training provided to school principals to improve their management capacities. This cost includes lodging, transportation, materials, facilities, and catering for beneficiaries. The capacity-building activities were designed around two training programs—one for the Stallings classroom observation and one for the SisAT tool. The implementation of the two training programs with professional trainers, which benefited the PEC Plus school principals, had an average cost of 472 USD per school—266 USD per school for the Stallings classroom observation tool and 206 USD per school for the SisAT tool.

Both of the training programs took place yearly in both years. Each program consisted of a five-day training session led by a team of professional trainers. The professional trainers included staff from the federal education authority (SEP) who had developed the tools.

Note that all schools participating in the impact evaluation (i.e., PEC, PEC Plus, and control), received the training in these tools via the cascade model. The training program was delivered to school supervisors by a team of professional trainers, who were subsequently responsible for training other school supervisors. Then, supervisors were responsible for training school principals. For PEC Plus schools, the professional team of trainers conducted additional training directly to school principals.

A.2 Short-term leadership certificate training program

Some of the school principals in PEC Plus schools received a short-term leadership certificate training program. The state-level education authorities selected this program based on Federal SEP guidelines related to national school principal's profile standards. These national school principal's profile standards define school principals as school and community leaders whose roles are the following: (i) know the school and classroom's dynamics, as well as the forms of organization and operation of the school; (ii) are recognized as professionals who continuously participate in professional development activities to improve the quality of the educational service; (iii) assume and promote the legal and ethical principles inherent to their roles and educational work, to ensure the right of students to quality education; and (iv) know the school's social and cultural context and establish collaborative relationships with the community, the school area, and other instances, to enrich the educational task.[25]

The DMW instrument used to measure school principals' managerial capacity is constructed around four dimensions of managerial practices, including operations management, people management, target setting, and monitoring. When adapting the DWS instrument to Mexico, the Federal SEP decided to include an additional dimension related to leadership as this aligned well with the national school principal's profile standards. However, the DWMS measure used to assess school principals' managerial capacity in this paper is only constructed using the original instrument's four dimensions.

[25]The school principals profile standards are available at: http://servicioprofesionaldocente.sep.gob.mx/portal-docente-2014-2018/2018/PPI_PROMOCION_EB_2018_19012018.pdf