Regular article

Direct vs indirect management training: Experimental evidence from schools in Mexico☆

Mauricio Romero [a,*], Juan Bedoya [b], Monica Yanez-Pagans [c], Marcela Silveyra [c], Rafael de Hoyos [a,c,d]

[a] *ITAM, Mexico*
[b] *Universidad de Cantabria, Spain*
[c] *The World Bank, United States of America*
[d] *Xaber, Mexico*

ARTICLE INFO

ABSTRACT

We use a large-scale randomized experiment (across 1,198 public primary schools in Mexico) to study the impact of providing schools directly with high-quality managerial training by professional trainers vis-à-vis through a cascade-style "train the trainer" model. The training focused on improving principals' capacities to collect and use data to monitor students' basic numeracy and literacy skills and to provide feedback to teachers on their instruction and pedagogical practices. After two years, the direct training improved schools' managerial capacity by 0.13σ (*p*-value 0.018) (relative to "train the trainer" schools), but had no meaningful impact on student test scores (we can rule out an effect greater than 0.08σ at the 95% level).

1. Introduction

Schools are complex organizations that are often poorly managed. Across developed and developing countries, they tend to have worse management practices than hospitals and manufacturing firms (Bloom et al., 2014, 2015a). This is not surprising; school principals are chosen according to seniority in many countries. As a result, although they have years of classroom experience, principals may lack management skills.

We study the implementation of the Government of Mexico's large-scale *Escuela al Centro* (in English, school at the center) strategy, designed to strengthen school autonomy and improve principals' managerial capacity. This strategy was implemented nationwide for three consecutive school years (2015–16, 2016–17, and 2017–18). A core component was managerial training for principals that focused on collecting and using data to monitor students' basic numeracy and literacy skills and providing feedback to teachers on their instruction and pedagogical practices. We randomly assigned 1,198 eligible public primary schools to one of two groups: (1) a "train the trainer" group, which received managerial training under a cascade model in which 10% of school supervisors were trained by professional trainers, who then trained other supervisors, who in turn provided training to principals (*n* = 599) and (2) a "direct training" group, in which principals received managerial training directly from a team of professional trainers (*n* = 599).[1]

---

[1] Supervisors are the direct link between schools and educational authorities in each state. Supervisors are typically in charge of 8 to 20 schools (Santiago et al., 2012).

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

We collected data on schools' managerial practices, using the Development World Management Survey (DWMS) (Lemos and Scur, 2016), at baseline (in late 2015) and two years after the program was implemented (in early 2018). The DWMS measures different dimensions of schools' managerial practices, including operations management, people management, target setting, and monitoring. To measure students' learning, we use data from a nationwide standardized test (PLANEA).[2]

Our results show a significant improvement of 0.13 (p-value 0.018) standard deviations (σ thereafter) in managerial capacities among "direct training" schools compared to "train the trainer" schools. The improvements in managerial capacities do not translate into meaningful impacts on student learning. Students in "direct training" schools have test scores that are 0.03σ higher than their counterparts in "train the trainer" schools. However, this difference is not statistically significant (p-value 0.24) and we can rule out an effect greater than 0.08σ on test scores at the 95% level. There is little evidence of heterogeneity in treatment effects by baseline school characteristics.

The failure of "direct training" to significantly 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 shows that a 1σ improvement in managerial practices is associated with an increase of less than 0.1σ in test scores, a weaker correlation than Bloom et al. (2015a) reported for several countries. However, even assuming a stronger link between management and test scores (an increase of 0.4σ in test scores as the management index increases by one standard deviation) based on the results from Bloom et al. (2015a) would imply that an increase of 0.13σ in management practices should yield an increase in test scores of 0.029σ—the actual treatment effect was 0.03σ.[3] Overall, the expected treatment effects on learning outcomes (assuming previous correlational evidence is causal and 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 statistically significant (even with a sample size of 1,198 schools) changes in learning outcomes. The fact we do not find treatment effects on test scores is not due to a lack of power. Our ex-post minimum detectable effect (MDE) is 0.081σ for test scores (with power of 80% and size of 5%) (Ioannidis et al., 2017; McKenzie and Ozier, 2019). Rather, this result likely implies the need for larger effects on management practices to find economically meaningful effects on test scores.[4]

One way to boost the intervention's impact on management practices would be to increase principals' attendance to the training workshops. While "direct training" principals were about ten percentage points more likely to complete courses or receive counseling on how to carry out school director duties in the past, less than 25% completed the entire training ($\sim$ 80 hours), and roughly 10% completed less than 20 hours of the training. Instrumental variable approaches suggest boosting attendance to the training workshops would result in further improvements in management that would translate into meaningful impacts in student learning outcomes. However, we take these results as suggestive evidence that requires further confirmation in future studies due to measurement errors in the attendance data.

We contribute to the literature and policy debate on improving school management in low- and middle-income countries. Our study

advances 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 differences in school effectiveness. However, 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 Fryer (2014)'s findings, Bloom et al. (2015a) document a positive and statistically significant correlation between managerial practices and student learning outcomes. There is also evidence from India that learning outcomes and progress are positively correlated with managerial practices (Lemos et al., 2021). Our baseline data adds to the evidence base on the correlation between school management and learning outcomes. We find a weaker correlation between them than previous studies have identified, which could be partially explained by the low autonomy in the Mexican public education system—Bloom et al. (2015a) shows higher school autonomy is correlated with higher management scores.[5]

Moreover, we provide experimental estimates of the relative effectiveness of two strategies to improve school principals' managerial capacity on management practices and student learning outcomes in a developing country. While there is evidence from the US that training programs to improve school principals' managerial practices have a positive effect on student learning outcomes (Fryer, 2017), our evidence and findings from other developing countries suggest otherwise. A closely related paper by Muralidharan and Singh (2020) shows that an attempt to improve management quality in Indian schools by inducing principals to adopt "best practices" had no impact on student outcomes. India's accountability and incentive structure for principals is rather weak (as it is in Mexico), which the authors argue may explain why improving managerial practices has little or no effect on test scores.[6]

2. Context and intervention

2.1. Context

Mexico's primary education system (grades 1 to 6) has more than 14 million students and 573,000 teachers distributed across roughly 100,000 schools.[7] The system is highly decentralized: 32 state-level education systems follow a common national curriculum and general guidelines from the Federal Secretariat of Public Education (Federal SEP, from its acronym in Spanish). However, local governments are fully responsible for administering each state-level Secretariat of Public Education.

Access to primary education in Mexico is high, with over 98% of children aged 6 to 12 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 basic proficiency in math and Spanish (and only one in three in marginalized areas) according to 2018 nationwide standardized tests (Instituto Nacional para la Evaluación de la Educación, 2018).

---

[3] Alternately, using our own data—and under some strong assumptions that allow us to use the treatment as an instrument for DWMS scores—our treatment effect on DWMS implies an expected increase of 0.065σ in test scores, given the treatment effect on DWMS scores.

[4] Alternatively, it could be the case that schools in Mexico are so well managed that the returns to additional increases in management are relatively low. However, comparing the distribution of DWMS scores in our setting to those in other countries found by Bloom et al. (2015a) suggests this is not the case.

[5] Mexican schools are less autonomous than schools in other Organization for Economic Co-operation and Development (OECD) countries (Hopkins et al., 2007; OECD, 2016).

[6] A second potential explanation for the lack of impact is that managerial practices take longer to improve student education outcomes (see de Hoyos et al., 2020).

[7] Unlike other countries in Latin America, Mexico has a small private education sector that accounts for only 10% of the total primary enrollment (Elacqua et al., 2018).

Mexico has three types of public primary schools: general primary schools (which teach most children), and indigenous and community schools, which serve roughly 800,000 and 400,000 students, respectively. These include many small, multi-grade schools with small numbers of students.[8] The existence of a large number of small schools increases the governance challenges and requires tailored management models.

These governance challenges are compounded by a high rotation of teachers and school principals and—until recently—the lack of a system to regulate the entry and promotion of teachers. Previously, the national teachers' union influenced teachers' (and school principals') appointments (Álvarez et al., 2007). In 2013, the central government implemented a major education reform that defined and regulated a merit-based process to hire and promote teachers and principals. It also introduced the *Escuela al Centro* strategy to enhance principals' managerial capacities to improve 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). It had two main component—the provision of school grants and school principals' managerial training.[9]

The grant component consisted of an annual monetary transfer to schools that submitted an improvement plan approved by their 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 improvement plans and pay for basic supplies and repairs. As explained in Section 3.1, all schools in our sample received these grants.

The training component focused on improving school principals' capacity to collect and use data to monitor students' basic numeracy and literacy skills and provide teachers with feedback on their teaching styles. To implement this training, the Federal SEP developed two tools: (i) a student assessment to monitor foundational 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 SisAT was developed based on evidence that providing school principals in Mexico with information on what areas of the national curriculum are the most challenging for students, based on national standardized learning assessments, had positive effects on student learning (de Hoyos et al., 2017, 2019). It includes items from past national standardized assessments to measure students' basic numeracy and literacy skills and identify lagging students to trigger early remedial actions. Teachers administer the SisAT and input the scores into a simple software program that generates a detailed report and flags students with significant learning gaps. The SisAT also pinpoints the most challenging areas of the national curriculum for students and classrooms. While schools were free to decide when to administer the SisAT, most did so at the beginning of the school year to generate baseline measures to include in their school improvement plans and throughout the school year to monitor students' progress.

The Stallings classroom observation tool was developed based on evidence that using school principals to coach teachers improves student learning in Mexico (Secretaría de Educación Pública and Banco Internacional de Reconstrucción y Fomento, 2015). It collects information on the teacher's use of time in the classroom, including the activities conducted, pedagogical practices, use of educational materials, and level of student engagement (Stallings, 1977; Stallings and Molhlman, 1988). The tool helps school principals systematically collect data to provide feedback to teachers on how to improve their teaching.

The Federal SEP developed a high-quality training strategy, including learning materials, to help principals use the SisAT and the Stallings classroom observation tool. The training consisted of 40 hours of instruction per tool.[10] The SEP used a "train the trainer" cascade model to roll out the *Escuela al Centro* strategy throughout the country. State-level education authorities selected 10% of all primary school supervisors to receive the training from a professional team that included staff involved in designing the tools. The trained supervisors then provided training to the other supervisors in their state. After all supervisors in a state were trained (by either the professionals or their peers), they then proceeded to train the school principals in their jurisdictions. To test the efficacy of the cascade model versus professional training, the SEP provided professional training to some school principals.

3. Research design and data

3.1. Sampling and randomization

To test the effectiveness of the professional training, the SEP invited all 32 states in the country to participate in an impact evaluation. The seven states that met the requirements—Durango, Estado de México, Morelos, Tlaxcala, Guanajuato, Tabasco, and Puebla—were selected to be part of this research study (see Figure A.1).[11]

The local education authorities invited all public primary schools in all seven states to apply for the school grant component of *Escuela al Centro*. We randomly assigned the 1,198 schools that applied to the grants to one of two groups: (1) "train the trainer" schools, which received a school grant and school principals' managerial training using the cascade model ($n = 599$) or (2) "direct training" schools, which received a school grant and professional training ($n = 599$).[12]

Our experimental design allows us to estimate the causal effects of using professional trainers vs. the cascade model to train school principals.[13]

Our sample included public primary schools that chose to participate in the program. To be eligible, schools had to have more than 60 students; those with at least one classroom with students from different grades were excluded.[14] Therefore, the schools included in the experiment have more students and teachers and are more likely to be urban than the average public primary school in Mexico (see Table A.1).

---

[8] The smallest 40% of primary schools in the country serve 8.5% of its primary school students. By comparison, Mexico has less than half of the student population of the United States, but 50% more schools.

[9] 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.

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

[11] From the 32 states in Mexico, 14 states expressed interest in participating in the impact evaluation. However, only seven complied with the required paperwork.

[12] Some principals in "direct training" schools also benefited from short-term leadership certificate training programs offered by state-level education authorities. These programs focused on leadership issues, in line with the national school principal's profile standards. As explained in more detail in Section 3.2, the DWMS—the instrument we use to measure principals' overall managerial practices—does not take leadership practices into account. Appendix A.5 provides further details on the states' short-term certification programs.

[13] While it is not possible to experimentally identify the impact of the cascade-style training vis-à-vis no training at all, there is evidence that cascade training models tend to be relatively ineffective (Popova et al., 2018).

[14] Small schools were excluded because the managerial intervention was focused on training principals to coach teachers. In small schools, principals also teach and thus need different management models.

The randomization protocol varied slightly across the seven participating states. Broadly, schools were first stratified into different groups (by enrollment and location) and then randomly assigned to either the treatment ("direct training" by professional trainers) or control (cascade-style training) group. Section A.3 details each state's sampling and randomization strategy.

3.2. Data

We collected primary data on the 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 number of teachers. Our study period coincides with two school years, 2015–16 (baseline) and 2017–18 (follow-up). In addition, the baseline and follow-up months roughly coincide with the nationwide standardized test application dates, which allow us to measure the intervention's impact on both management practices and student test scores.

3.2.1. Primary data

Information on schools' managerial practices was collected using the DWMS—an adaptation of the World Management Survey (WMS), originally developed to measure the quality of management practices in manufacturing firms in developed (Bloom and Van Reenen, 2007) and developing countries (Bloom et al., 2013).[15] The WMS and the DWMS were subsequently adapted to measure management quality in the education and health sectors (Bloom et al., 2015a,b). 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 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.[16] The interviews, conducted by a team of two trained enumerators (one coder and one interviewer), 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. We assigned the same team of trained enumerators to code the audio files from all the original interviews to ensure comparability over time. Unfortunately, 32% of the audio files from the baseline, and 16% from the follow-up, were damaged when 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 Tables A.2 and A.3). Thus, our results are unlikely to be driven by differences in observable or unobservable characteristics between schools with and without functioning audio files, including the treatment status. To ensure comparability across schools, we randomly assigned audio files to enumerators and control for enumerator fixed effects in all the regressions. We conducted the baseline DWMS surveys between October 2015 and May 2016, and the follow-up surveys from January to May 2018.[17]

For reference, we compare the distribution of management scores in our setting (at baseline) to the distribution in India, Brazil, and the US from Bloom et al. (2015a)—see Figure A.2. Overall, the average school in our setting has a higher management score than the average school in India (2.1 vs 1.7), a similar score to the average school in Brazil (2.1 vs 2.0) and a lower score than the average school in the US (2.1 vs 2.7). However, the dispersion in management practices in our setting is lower, which could be explained by the restrictions imposed on the experimental sample (e.g., excluding small multi-grade public schools and all private schools).

School principals also completed two online surveys to assess the quality of managerial training—one for each tool. The surveys included questions about different elements of the tools and their associated training. Since the surveys were not mandatory, many school principals did not complete them. Schools that answered the online surveys are statistically different from those that did not in several observable characteristics, including the treatment status (see Tables A.9–A.12). For completeness, we report some basic statistics from these two online surveys. However, their information is not representative of our experimental sample due to sample selection (i.e., it has differential attrition across treatments and within each treatment); therefore, we exclude this data from our main analysis.

3.2.2. Secondary data

We use three types of secondary data. First, we measure student learning outcomes using PLANEA test scores. The exam was administered to grade 6 in June 2015 and June 2018. SEP gave the authors access to anonymized student-level data for both years for all schools in our sample. As part of registering their school for PLANEA, principals needed to fill a survey (PLANEA-Contexto). The survey asks about their daily activities and the challenges they face. We use these surveys as a secondary measure of principals' management practices and their exposure to the training.

Second, we gathered information on the location of each school from the PLANEA data. We use this information to match each school to its locality's marginalization index, which accounts for deficiencies in education, housing, population, and household income.[18] Third, we use administrative school census data collected by 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, the school principal's characteristics, the number of classrooms, and its geographic location. This school census data can be matched with the PLANEA data.[19]

3.3. Balance and attrition

Most student and school characteristics are balanced across treatment arms at baseline (see Table 1). The average school in our sample has 279 students, 9.4 teachers, and a pupil–teacher ratio of 29; 40% of schools are in rural areas and 38% 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 endline DWMS and PLANEA data (in 2018). We have PLANEA data for nearly all schools (~99%) and DWMS data for ~77% of schools (due to damaged audio files from the interviews, as mentioned above). The proportion of schools with both PLANEA and DWMS data is balanced across treatments.

---

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

[16] The DWMS adaptation for Mexico included an additional dimension, leadership. Having this additional dimension responded to the government's need to better align the DWMS instrument to the rules of operation of *Escuela al Centro*. All the analyses reported in this paper exclude the leadership dimension when constructing the overall DMWS index to ensure it is comparable with other settings.

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

[18] *Consejo Nacional de Población* (CONAPO) estimates this index.

[19] All the data used in this paper can be downloaded from www.xaber.org.mx.

Table 1
Balance across treatment groups.

	(1) Mean (SD)	(2)	(3) Difference
	Train the trainer	Direct training	(2)-(1)
Students in math achievement L-IV (%)	7.79	8.36	0.56
	(11.11)	(12.25)	(0.66)
Students in math achievement L-I (%)	60.00	60.17	0.19
	(21.81)	(22.24)	(1.22)
Students in language achievement L-IV (%)	2.67	3.31	0.65**
	(3.86)	(6.40)	(0.30)
Students in language achievement L-I (%)	52.17	51.56	−0.60
	(20.25)	(20.52)	(1.15)
Marginalization	0.38	0.38	−0.00
	(0.49)	(0.49)	(0.02)
Urbanization	0.41	0.39	−0.02
	(0.49)	(0.49)	(0.02)
Number of students	272.59	285.96	13.31
	(163.74)	(163.69)	(8.87)
Number of teachers	9.27	9.63	0.36
	(4.23)	(4.39)	(0.24)
Student–teacher ratio	28.34	28.89	0.54
	(6.92)	(7.18)	(0.35)
DWMS endline missing	0.22	0.23	0.01
	(0.41)	(0.42)	(0.02)
PLANEA endline missing	0.01	0.01	0.00
	(0.08)	(0.09)	(0.00)
Observations	599	599	1,197

This table presents the means and standard deviations (in parentheses) for "train the trainer" (Column 1) and "direct training" schools (Column 2). The differences reported in Column 3 take into account the randomization design (i.e., including strata fixed effects), and standard errors (in parentheses) are clustered at the school level. Achievement level (L) refers to the PLANEA 2015 exam results, which are scored from L-I (lowest) to L-IV (highest). *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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.4. Compliance

To measure compliance with the evaluation's original design, we compiled information on whether school principals reported attending the training sessions on the two tools. As mentioned above, since the characteristics of schools that answered the survey are different from those that did not (see Tables A.9–A.12), these results should be interpreted with caution. Due to the sample selection in the compliance measures and the inability to directly compare the training hours across treatment arms (cascade vs. direct), local average treatment effect estimates using the treatment assignment as an instrument for the number of training hours principals report are difficult to interpret and likely biased.

While virtually no principals in "train the trainer" schools completed the full training on the use of either tool, less than half (∼40%) received some training (10–39 hours) through the cascade model (see Columns 1 and 2 of Table 2). About one-quarter of principals in "direct training" schools (20%–25%) completed the training on both tools, and roughly 80% received some training from professionals. The difference between treatment groups is statistically significant for both completed training and the indicator for some training. This is further supported by evidence from surveys principals completed as part of the nationwide student standardized test (PLANEA-Contexto surveys) in 2018. Specifically, "direct training" principals were more likely to complete courses or receive counseling on how to carry out school director duties in the past 12 months (see Panel A, Table A.4).

4. Results

4.1. Correlation between management (DWMS) and learning

We first explore the correlation between learning outcomes and DWMS at baseline. We seek to replicate the analysis in Bloom et al.

Table 2
Compliance across treatment groups.

	(1) Mean (SD)	(2)	(3) Difference
	Train the trainer	Direct training	(2)-(1)
Panel A: Stallings classroom observation tool			
All training sessions (40 hours)	0.01	0.24	0.23***
	(0.10)	(0.43)	(0.02)
Some training sessions (10-40 hours)	0.39	0.86	0.44***
	(0.49)	(0.35)	(0.03)
Observations	304	533	837
Panel B: Foundational skills measurement tool (SisAT)			
All training sessions (40 hours)	0.01	0.19	0.18***
	(0.09)	(0.39)	(0.02)
Some training sessions (10-40 hours)	0.32	0.72	0.39***
	(0.47)	(0.45)	(0.03)
Observations	402	464	866

This table presents the means and standard deviations (in parentheses) for "train the trainer" (Column 1) and "direct training" schools (Column 2). The differences reported in Column 3 take into account the randomization design (i.e., including strata fixed effects), 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). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(2015a) 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 3). 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 et al. (2015a) and control for the number of pupils in the school, the pupil–teacher ratio, and the marginalization index (Column

Table 3
Association between DWMS and test scores at baseline (all schools in the sample).

	(1)	(2)	(3)	(4)	(5)
	PLANEA 2015 scores				
DWMS	0.0017	0.011	0.020	0.017	−0.0065
	(0.025)	(0.025)	(0.023)	(0.022)	(0.027)
No. of obs.	20,680	20,680	20,680	20,049	20,049
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). 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 et al. (2015a) 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 0.104σ. 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 was the most closely correlated with student outcomes, followed by monitoring and people; none of them demonstrated a statistically significant correlation with test scores in our setting (see Table A.5).

4.2. Experimental results

Our main estimating equation for student-level outcomes is:

$$Y_{isg} = \alpha_g + \gamma_1 DirectTraining_s + \varepsilon_{isg} \qquad (1)$$

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, $DirectTraining_s$ indicates whether school s received training directly provided by professional trainers, 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 standard errors at the school level. γ_1 is the coefficient of interest and reflects the difference between the two types of training.

Overall, the direct training intervention improved management practices relative to the indirect training (see Panel A, Table 4). Management scores in schools that received direct training were 0.13σ (p-value 0.018) higher than in "train the trainer" schools. Therefore, our results show that it pays off to invest in professional trainers to improve school principals' management capacities.[20]

---

[20] According to surveys administered to principals as part of the nationwide student standardized test (PLANEA-Contexto surveys), in 2018 "direct training" principals were not more likely than those trained using the cascade method to undertake activities to improve learning outcomes, observe classroom teaching, help teachers improve their pedagogical practices, or provide parents with school and student performance information (see Panel B, Table A.4). However, these self-reported measures are likely inflated by social desirability bias given the (likely unrealistic) high proportion of principals who report doing these activities often or very often. Thus, we do not believe the difference between the "train the trainer" and "direct training" from these self-reported measures accurately reflects treatment effects.

Table 4
Effects on the DWMS and on learning outcomes.

Panel A: DWMS and its components						
	(1)	(2)	(3)	(4)	(5)	(6)
	DWMS	Operations	Monitoring	Targets	People	Leadership
Direct training	0.13**	0.14**	0.13**	0.027	0.093*	−0.0091
	(0.053)	(0.056)	(0.060)	(0.052)	(0.056)	(0.060)
No. of obs.	913	913	913	913	913	911

Panel B: Learning outcomes				
	(1)	(2)	(3)	(4)
	Math	Language	Average	PCA
Direct training	0.031	0.027	0.035	0.035
	(0.029)	(0.027)	(0.029)	(0.029)
No. of obs.	39,263	39,665	37,958	37,958

Panel A 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. Finally, Column 6 has the additional dimension, leadership; the SEP asked for this dimension to be measure in addition to the four traditional components of the DWMS. The overall DMWS index used in Column 1 excludes the leadership dimension to ensure comparability with other settings. Panel B 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 account for the randomization design (i.e., they include strata fixed effects). Panel A regressions also include enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Given the nature of the intervention (direct vs. indirect training on the Stallings and the SisAT tools) is not surprising that the "Operations" and "Monitoring" dimensions improve the most. "Operations" partially measures whether there is data-driven planning, as well as personalization of instruction and learning—goals the Stallings and the SisAT specifically help with. Likewise, "Monitoring" partially measures whether school performance is measured frequently and appropriately (SisAT does this for students, and Stallings does it for teachers). Given the limitations principals face to dismiss or promote teachers, it is not surprising that the treatment effect on "People/talent management" is lower. However, measuring teachers' performance (via Stallings) enables principals to provide soft incentives (e.g., better teaching assignments or non-pecuniary rewards).[21]

While management practices improved as a result of the direct training intervention, test scores did not (see Panel B, Table 4). Students in "direct training" schools scored 0.03σ (p-value 0.24) higher than those in "train the trainer" schools. We can rule out, at the 95% confidence level, the possibility that test scores increased by more than 0.09σ with respect to "train the trainer" schools. This result is robust to a series of student- and school-level controls (see Table A.6). Including controls allows us to rule out an effect greater than 0.08σ at the 95% level. Finally, there is no evidence that the "direct training" affected other outcomes such as grade repetition or enrollment rates (see Table A.8).

4.3. Discussion: The lack of effect of direct training on test scores

As mentioned above, Bloom et al. (2015a) find that a one-standard-deviation increase in the WMS index is associated with an increase in

---

[21] We further explore whether it is reasonable to expect that providing training on two tools would improve managerial practices in Section A.2. We address this question by looking at the correlation between the self-reported information on the use of the Stallings classroom observation and SisAT tools on both DWMS. We find that "direct training" schools are more likely to use the management tools provided to them, and the use of these tools is correlated with the DWMS. However, since schools that answered these surveys are statistically different from those that did not in several observable characteristics, including treatment status (see Tables A.9–A.12), these correlations may be biased and are presented for completeness.

pupil outcomes of 0.2σ–0.4σ. The evidence from our baseline shows a weaker correlation between management practices and test scores. Thus, optimistically assuming that a one-standard-deviation increase in management practices generates a treatment effect of 0.4σ on student learning, an increase of 0.13σ in management practices should yield an increase in test scores of 0.029σ—the actual treatment effect was 0.03σ.

We also estimate the effect of an increase in the DWMS index on test scores using the treatment assignment to instrument for the DWMS index. While this requires a strong assumption that the DWMS completely captures any possible effect of the treatment on test scores, it provides a different benchmark of the plausible causal effect of improvements in management practices on test scores. The instrumental variable approach suggests increasing the DWMS by one standard deviation increases test scores by 0.49σ (see Table A.7). This implies an expected increase of 0.065σ in test scores, given the treatment effect on DWMS scores.

Further, the components of the DWMS index that Bloom et al. (2015a) find are more associated with test scores, are the ones where the direct training intervention improved management practices the least relative to the indirect training (see Columns 2–4 in Panel A of Table 4). Specifically, the treatment effect on the two components that have the highest association with learning outcomes ("people/latent management" and "target setting") are the lowest.[22]

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 direct training intervention improved management practices relative to the indirect training, these improvements did not generate statistically significant changes in learning outcomes (even with a sample size of 1,198 schools). However, we cannot rule out the possibility that management had a *small* positive impact on learning.

Given the low overall attendance rate to the training workshops (see Section 3.4), we explore whether increasing participation in the training workshops would result in further improvements in management practices and larger learning gains. To answer this question we use an instrumental variable approach to study the effects of attending more training workshops. Specifically, we instrument attendance to training workshops with whether a school was randomly assigned to "direct training".

However, we face a trade-off between two different approaches to measure workshop attendance. We could use PLANEA-Contexto surveys, which all principals answered, but that do not ask about training workshops from our program specifically, but rather about any courses or counseling on how to carry out school director duties in the past. On the other hand, using the online surveys to measure (self-reported) attendance to the training workshops in this program will likely induce sample selection bias since the characteristics of schools that answered the survey are different from those that did not. We report both. While neither approach is perfect, both suggest similar results.

The PLANEA-Contexto surveys suggest that attending any courses or counseling on how to carry out school director duties increases both management practices and learning outcomes (see Panel B, Table 5). The local average treatment effects (LATE) here represent the effects of attending any workshops, not just those related to our program, for the compliers who are more likely to attend a workshop due to the "direct training" treatment. While attending any courses or counseling on how

Table 5
Effects of principal's attendance to the training workshops.

	(1) DWMS	(2) DWMS	(3) PLANEA	(4) PLANEA
Panel A: Online surveys				
Attended > 10 hours of training	0.36***		0.15*	
	(0.11)		(0.070)	
Attended all trainings		0.69**		0.28*
		(0.21)		(0.13)
N. of obs.	808	808	28,906	28,906
F test (first stage)	292	143	240	138
Panel B: PLANEA - Contexto				
Ever	1.1*		.68*	
	(.55)		(.39)	
Past 12 months		1**		.56*
		(.5)		(.31)
N. of obs.	850	850	29,731	29,731
F test (first stage)	16	26	13	30

Panel A presents the effect of a principal attending at least 10 hours of training on the DWMS score (Columns 1) and the overall PLANEA score (Column 3), as well as the effect of a principal attending all training on the DWMS score (Columns 2) and the overall PLANEA score (Column 4). Attendance (in both cases) is instrumented with the treatment allocation. The F statistic of the first stage is presented in the bottom row (see Table 2 for details on the first stage). Columns 1–2 use data at the school level, while Columns 3–4 use data at the student level. Attendance is measured using online surveys which have differential attrition across treatments (see Tables A.9–A.12). Panel B presents the effect of a principal ever attending a training workshop (on any topic related to his or her duties) on the DWMS score (Columns 1) and the overall PLANEA score (Column 3), as well as the effect of a principal attending a training workshop (on any topic related to his or her duties) in the past 12 months on the DWMS score (Columns 2) and the overall PLANEA score (Column 4). Attendance (in both cases) is instrumented with the treatment allocation. The F statistic of the first stage is presented in the bottom row (see Table A.4 for details on the first stage). Columns 1–2 use data at the school level, while Columns 3–4 use data at the student level. Attendance is measured using PLANEA-Contexto surveys which do not have differential attrition across treatments (see Table 1). All regressions account for the randomization design (i.e., they include strata fixed effects) and include enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to carry out school director duties is likely to capture a significant portion of the effect of "direct training", it is unlikely to be the only channel through which the treatment affects outcomes—a necessary condition for the LATE to be valid.

The online surveys suggest that attending the training workshops from this program increases both management practices and learning outcomes (see Panel A, Table 5). However, the LATEs are likely biased due to sample selection caused by the differential attrition in the survey. In addition, and as mentioned above, the training hours across the treatment arms (cascade vs. direct) are not directly comparable.

Overall, while both approaches have limitations, they suggest one way to boost the intervention's impact on management practices and learning outcomes would be to increase principals' attendance to the training workshops.

4.4. Heterogeneity

Next, we explore heterogeneous treatment effects on management practices by schools' (and principals') baseline characteristics. Overall, there is little evidence of heterogeneity. Specifically, we estimate the following equation:

$$Y_{isg} = \alpha_g + \beta_1 treatment_s + \beta_2 treatment_s \times c_s + \beta_3 c_s + \varepsilon_{isg} \quad (2)$$

where c_s denotes the school characteristics of 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 Eq. (1). We study heterogeneity in schools' baseline management quality, marginalization index, and principals' gender and tenure. Overall, we find no evidence of heterogeneity in management practices (DWMS) or learning outcomes (see Tables A.13 and A.14).

---

[22] Bloom et al. (2015a) find that of the four components of the DWMS, "people/latent management" had the highest association with test scores (an increase of one standard deviation in "people/latent management" score was associated with an increase of 0.257 standard deviation in pupil test scores), followed by "target setting" (associated with an increase of 0.158 standard deviation in pupil test scores), "monitoring" (associated with an increase of 0.133 standard deviation in pupil test scores) and "operations" (associated with an increase of 0.093 standard deviation in pupil test scores).

We also study whether there is heterogeneity by whether there was a change in the school's principal between 2015 and 2018. We first assess that the treatment did not have an impact on principal turnover itself (see Table A.16), but note that ∼43% of schools change principals at some point in those three years. While high principal turnover may be a barrier to improving learning outcomes (Miller, 2013; Bartanen et al., 2019), there is no heterogeneity in treatment effects on management practices or learning outcomes by teacher turnover (see Table A.16).

5. Conclusions

Recent studies have identified the pivotal role that managerial practices play in helping an organization achieve its objectives (Bender et al., 2018), and the education sector is no exception. This paper reports some of the first experimental evidence of the relative effectiveness of two interventions to improve school management in a developing country. We randomly assigned a group of public primary schools in seven Mexican states to receive training either directly from professional trainers or a "train the trainer" cascade model. Compared to indirect training, direct training improved school principals' managerial capacity but failed to improve learning outcomes. To improve student learning in the short term, a management intervention may need to have a greater impact on school principals' managerial capacities.

However, given the cost of the "direct training" intervention (∼470 USD per school, see Appendix A.4), the marginal dollar in Mexico might be better spent on interventions that focus on improving pedagogy (e.g., teaching at the right level, teacher content and pedagogical training) and improving teacher accountability (Kremer et al., 2013; Glewwe and Muralidharan, 2016; Snilstveit et al., 2016).

CRediT authorship contribution statement

Mauricio Romero: Conceptualization, Data curation, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Juan Bedoya:** Software, Data curation, Visualization, Formal analysis. **Monica Yanez-Pagans:** Project administration, Data curation, Writing – review & editing, Funding acquisition. **Marcela Silveyra:** Project administration, Data curation, Writing – review & editing, Funding acquisition, Resources. **Rafael de Hoyos:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2021.102779.

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.

Bartanen, B., Grissom, J.A., Rogers, L.K., 2019. The impacts of principal turnover. Educ. Eval. Policy Anal. 41 (3), 350–374. http://dx.doi.org/10.3102/0162373719855044.

Bender, S., Bloom, N., Card, D., Van Reenen, J., Wolter, S., 2018. Management practices, workforce selection, and productivity. J. Labor Econ. 36 (S1), S371–S409. http://dx.doi.org/10.1086/694107.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., Roberts, J., 2013. Does management matter? Evidence from India. Q. J. Econ. 128 (1), 1–51.

Bloom, N., Lemos, R., Sadun, R., Scur, D., Van Reenen, J., 2014. The new empirical economics of management. J. Eur. Econom. Assoc. 12 (4), 835–876.

Bloom, N., Lemos, R., Sadun, R., Van Reenen, J., 2015a. Does management matter in schools? Econ. J. 125 (584), 647–674.

Bloom, N., Propper, C., Seiler, S., Van Reenen, J., 2015b. The impact of competition on management quality: Evidence from public hospitals. Rev. Econom. Stud. 82 (2), 457–489.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. Q. J. Econ. 122 (4), 1351–1408. http://dx.doi.org/10.1162/qjec.2007.122.4.1351.

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, https://www.planeacion.sep.gob.mx/estadisticaeindicadores.aspx.

Dobbie, W., Fryer, R., 2013. Getting beneath the Veil of Effective Schools: Evidence from New York City. Am. Econ. J. Appl. Econ. 5 (4), 28–60. http://dx.doi.org/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, http://dx.doi.org/10.18235/0001394.

Fryer, R., 2014. Injecting charter school best practices into traditional public schools: Evidence from field experiments. Q. J. Econ. 129 (3), 1355–1407.

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

Glewwe, P., Muralidharan, K., 2016. Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In: Eric A. Hanushek, S.M., Woessmann, L. (Eds.), In: Handbook of the Economics of Education, vol. 5, Elsevier, pp. 653–743, (Chapter 10).

Hopkins, D., Ahtaridou, E., Matthews, P., Posner, C., Figueroa, D., 2007. Reflections on the performance of the Mexican education system. OCDE. Directorate for Education, Disponible En www.Sep.Gob.Mx/Work/Models/Sep1/Resource/93128/5/Mex_PISA-OCDE2006.Pdf.

de Hoyos, R., Ganimian, A.J., Holland, P.A., 2019. Teaching with the test: Experimental evidence on diagnostic feedback and capacity building for public schools in Argentina. World Bank Econ. Rev..

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. Econ. Educ. Rev. 58, 123–140.

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

Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The power of bias in economics research. Econ. J. 127 (605), F236–F265. http://dx.doi.org/10.1111/ecoj.12461, https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12461.

Kremer, M., Brannen, C., Glennerster, R., 2013. The challenge of education and learning in the developing world. Science 340 (6130), 297–300.

Lemos, R., Muralidharan, K., Scur, D., 2021. Personnel management and school productivity: Evidence from India. Working Paper Series No. 28336. National Bureau of Economic Research, http://dx.doi.org/10.3386/w28336. http://www.nber.org/papers/w28336.

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

McKenzie, D., Ozier, O., 2019. Why ex-post power using estimated effect sizes is bad, but an ex-post MDE is not. World Bank Development Impact Blog.

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

Muralidharan, K., Singh, A., 2020. Improving public sector management at scale? Experimental evidence on school governance India. Working Paper Series No. 28129. National Bureau of Economic Research, http://dx.doi.org/10.3386/w28129. http://www.nber.org/papers/w28129.

OECD, 2016. PISA 2015 Results (Volume II): Policies and Practices for Successful Schools. OECD Publishing.

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.

Santiago, P., McGregor, I., Nusche, D., Ravela, P., Toledo, D., 2012. OECD Reviews of Evaluation and Assessment in Education: Mexico 2012. OECD Publishing, http://dx.doi.org/10.1787/9789264172647-en.

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., http://escuelaalcentro.com/wp-content/uploads/2018/04/Documento-base-Evaluaci%C3%B3n-de-Impacto.pdf.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., Jobse, H., Schmidt, T., Jimenez, E., 2016. The impact of education programmes on learning and school participation in low-and middle-income countries.

Stallings, J., 1977. Learning to Look: a Handbook on Classroom Observation and Teaching Models. In: Wadsworth Series in Curriculum and Instruction, Wadsworth Pub. Co..

Stallings, J., Molhman, G., 1988. Classroom observation techniques. In: Keeves, J. (Ed.), Educational Research, Methodology and Measurement: An International Handbook. Elsevier Science & Technology Books.

World Bank, 2007. What is school-based management? Working Paper Series No. 44922. http://documents.worldbank.org/curated/en/113901468140944134/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.

Further reading

References cited in Supplementary material

Bruns, B., Luque, J., 2014. Great Teachers: How to Raise Student Learning In Latin America and the Caribbean. World Bank Publications.

INEGI, 2018. Marco geoestadístico de Mexico. https://www.inegi.org.mx/temas/mg/default.html#. (Accessed 6 January 2018).