ONLINE APPENDICES

Appendix A: Formal Model

Appendix B: Empirical Setting — Key Organizations and Individuals

Appendix C: Managerial Engagement Index Construction

Appendix D: Additional Data Descriptives and Visualizations

Appendix E: Robustness Checks and Alternative Specifications

Online Appendix A

In this Online Appendix, we mathematically analyze the key effects proposed by our hypotheses. Suppose there is an agreement in which a consulting partner is promoting the adoption of managerial practices by a recipient organization. Consider also that the latter organization has a manager responsible for translating and facilitating the implementation of these practices to several units. In our specific empirical context, the units are schools, whereas we refer to this manager as a middle manager—a middle-level public employee who heads a district of schools and is in direct touch with the consulting partner (a nonprofit specialized in promoting managerial practices in schools).

Some of those units are originally selected to implement the managerial practices promoted by the consulting partner. The sets of targeted and non-targeted units are expressed by the indicator $T_i \in \{0,1\}$, where $T_i = 1$ if unit i is targeted, 0 otherwise. The consulting partner chooses a level of effort e_c to promote the implementation of practices to the targeted units, such that $e_c > 0$ if $T_i = 1$ and $e_c = 0$ if $T_i = 0$. The manager, in turn, employs a level of effort e_f to oversee the activities of all units. We assume that this choice of effort is fixed, that is, $e_f = \bar{e_f} > 0$ (for instance, it derives from bureaucratic requirements that the manager must follow). Importantly, the manager can also learn from the consulting partner and facilitate the implementation of the same practices by either the targeted or *non-targeted* units. Denote as e_1 the dissemination effort that the manager dedicates to the group of targeted units and e_0 as the dissemination effort dedicated to the group of non-targeted units. As such, the total dissemination effort of the manager is then given by $e_m = \bar{e_f} + e_1 + e_0$.

We assume that the performance of each unit i at period t, denoted as $y_{i,t}$, is a function of the implementation of managerial practices. Such a performance-enhancing effect derives from the effort

exerted by the consulting partner (e_c) or the recipient organization manager (e_m). We thus specify the performance of units for each group of targeted and non-targeted units respectively as:

$$y_{i,t}(T_i = 1) = f(e_c + ME * (\bar{e_f} + e_1)), \text{ and}$$

$$y_{i,t}(T_i = 0) = f(ME * (\bar{e_f} + e_0)),$$

where $f(.)$ is a strictly concave function with respect to effort, that is, $df/de_. > 0$ and $d^2f/de_.^2 < 0$, where $e_.$ generically denotes all types of effort. The assumption of strict concavity is justified, given that we expect diminishing returns for efforts concerning performance. We also normalize $f(0) = 0$.

In turn, ME represents the manager's engagement in disseminating practices: the extent to which the manager crafts the necessary routines and builds the working relationships needed to internalize and disseminate the practices introduced by the consulting partner, in either the targeted or the non-targeted units. For simplicity, we compare a case involving positive managerial engagement (i.e., $ME = ME^{\text{High}} > 0$), to a case where managerial engagement is absent (i.e., $ME = ME^{\text{Low}}$, assumed to be equal to 0). In the latter case, the manager is unable to affect the performance of organizational units via her/his individual dissemination effort.

We also assume that the potential (ex-ante expected performance) of all units is the same before the implementation of managerial practices, that is, $E[y_{i,t-1} \mid T_i = 1] = E[y_{i,t-1} \mid T_i = 0]$. In our empirical context, as detailed in our methodology, this assumption is quite plausible since the targeted units were selected through a random sorting after targeted and non-targeted units were matched based on key observable traits. Considering these assumptions, we state our first proposition:

Proposition 1. If the recipient manager applies equal effort in the targeted and non-targeted units ($e_0 = e_1$), the presence of managerial engagement ($ME = ME^{\text{High}} > ME^{\text{Low}} = 0$) reduces the performance gap between targeted and non-targeted units.

Proof. Assuming $f(0) = 0$, if $ME = ME^{\text{Low}} = 0$, then $y_{i,t}(T_i = 1) - y_{i,t}(T_i = 0) = f(e_c)$. Meanwhile, considering $e_0 = e_1 = \bar{e}$, if $ME = ME^{\text{High}} > 0$, then $y_{i,t}(T_i = 1) - y_{i,t}(T_i = 0) = f\left(e_c + ME^{\text{High}} * (\bar{e_f} + \bar{e})\right) - f\left(ME^{\text{High}} * (\bar{e_f} + \bar{e})\right)$. As $f()$ is a strictly concave function, we know $f(x) > f(x + y) - f(y), \forall\, x, y > 0$. Given $e_c, ME^{\text{High}} * (\bar{e_f} + \bar{e}) > 0$, it follows that $f(e_c) > f\left(e_c + ME^{\text{High}} * (\bar{e_f} + \bar{e})\right) - f\left(ME^{\text{High}} * (\bar{e_f} + \bar{e})\right)$. Finally, this is the same as:

$$(y_{i,t}(T_i = 1) - y_{i,t}(T_i = 0))|ME = ME^{Low} > (y_{i,t}(T_i = 1) - y_{i,t}(T_i = 0))|ME = ME^{High},$$

showing a reduced performance gap due to the effects of ME^{High} reducing the marginal effect of the consulting partner on targeted units. Notice that it also holds for $e_0 = e_1 = 0$, given that $\bar{e_f} > 0$. ■

Intuitively, because an engaged manager can act on the targeted and non-targeted units by disseminating performance-enhancing managerial practices (given her positive level of managerial engagement, $ME = ME^{\text{High}} > 0$), she/he ends up increasing the overall performance of units and diminishing the marginal effect of the consulting partner.

Yet, the manager may not symmetrically exert equal effort in the targeted and non-targeted units. Suppose that the manager's utility function is given by:

$$U = \gamma * (Y^{T=1}(ME, e_c, \bar{e_f}, e_1) + Y^{T=0}(ME, \bar{e_f}, e_0)) + \omega - C(\bar{e_f} + e_1 + e_0),$$

where $Y^{T=1}, Y^{T=0}$ represent the aggregated performance of targeted or non-targeted units (i.e., $Y^{T=1} = \sum y_i(ME, e_c, \bar{e_f}, e_1)\,|T_i = 1$ and $Y^{T=0} = \sum y_i(ME, \bar{e_f}, e_0)\,|T_i = 0$, such that $Y^{T=1} + Y^{T=0} = Y(ME, e_c, \bar{e_f}, e_1, e_0)$ is the total aggregated performance conditional on effort choices). In turn, $\omega > 0$ is the manager's (fixed) wage, while C represents managerial costs as a function of effort, assumed to be an increasing and strictly convex function (i.e., $dC/de_{\cdot} > 0$ and $d^2C/de_{\cdot}^2 > 0$).

Parameter γ ($\gamma > 0$) is crucial and indicates that the manager derives utility from the *aggregated* performance of the targeted and non-targeted units. For example, the manager might individually value the aggregated performance of her units due to professional norms of conduct. Alternatively, she might also be subject to performance-contingent payments conditional on the outcomes of both the targeted and the non-targeted units. Both mechanisms would lead to higher levels of γ.

We, therefore, arrive at our central proposition that asymmetric managerial effort towards the non-targeted units will further increase the performance of non-targeted units, reducing the performance gap between targeted and non-targeted units generated by the consulting partner's efforts:

Proposition 2. Managerial engagement ($ME = ME^{\text{High}} > ME^{\text{Low}} = 0$) would induce managers to dedicate higher dissemination effort to the non-targeted units than to the targeted units ($e_0^* > e_1^*$). This effect further reduces the performance gap between targeted and non-targeted units.

Proof. Given that the manager maximizes $U(e_1, e_0)$ subject to e_c, e_f, and ME, considering $e_c, e_f, ME = ME^{\text{High}} > 0$, she/he will optimally choose $(e_0, e_1) = (e_0^*, e_1^*)$ such that the marginal benefits of efforts equal their marginal costs[1]. At the optimal level, we have $dy_i(T_i = 0)/de_0 = dC/de_0$ and $dy_i(T_i = 1)/de_1 = dC/de_1$. By assumption and following the chain rule, $dy_i/de = dy_i/df * df/de > 0$. As such, *y* reflects the concavity of *f*. Then, as y_i is strictly concave, $e_c > 0$ implies that $dy_i(T_i = 0)/de_0 > dy_i(T_i = 1)/de_1$. Given that γ equally weights the aggregated performances ($Y^{T=1}, Y^{T=0}$), *C* is strictly convex with respect to total efforts, and ex-ante performance of targeted and non-targeted units are equal by assumption, the manager will focus on the non-targeted units as these provide her/him with higher marginal benefit. It follows that asymmetrical levels of effort $e_0^* > e_1^*$ increase the manager's utility. Further, following our previous assumptions, given that partner efforts, manager total efforts and everything else is constant (i.e., $e_c = \overline{e_c}$ and $\sum e_m(i) = e_f + e_1 + e_0 = \overline{e_m}$), it is straightforward that:

$$\left(Y^{T=1} - Y^{T=0} \middle| e_c = \overline{e_c}, e_f = \overline{e_f}, ME = ME^{\text{High}} > 0, e_0^* > e_1^*\right) <$$
$$\left(Y^{T=1} - Y^{T=0} \middle| e_c = \overline{e_c}, e_f = \overline{e_f}, ME = ME^{\text{High}} > 0, e_0 = e_1\right).\blacksquare$$

Intuitively, as the targeted units already received training directly from the consulting partner, our concavity assumption (diminishing returns of effort) implies that the manager will derive greater marginal benefits by acting on the non-targeted units. Proposition 2 provides the theoretical basis of Hypothesis 3 presented in the main text. Although managerial engagement can result in generalized performance improvements in both the targeted and non-targeted units (as predicted by H2), the possibility that the manager may devote greater effort to the latter further reduces the performance gap between the two groups (H3). Econometrically, this effect can be investigated through a difference-in-differences methodology, explained in the main text (equation 1). We essentially observed how the average effect of the consulting partner in enhancing performance evolution, $E[\Delta y_{i,t} \mid T_i = 1] - E[\Delta y_{i,t} \mid T_i = 0]$, is influenced by the presence of managers with greater engagement.

[1] If e_c is sufficiently high, then we might have a corner solution with $e_c > e_0^* > e_1^* = 0$. Otherwise, $e_1^* > 0$ only if e_c is sufficiently low such that its performance-enhancing effect would not suffice to surpass the effect of ME $*$ $(e_0^* + \overline{e_f})$, and then the manager would optimally employ efforts $(e_0, e_1) = (e_0^*, e_1^*)$ such that ME $*$ $(e_0^* + \overline{e_f}) =$ ME $*$ $(e_1^* + \overline{e_f}) + e_c$, perfectly balancing $Y^{T=1} = Y^{T=0}$.

Interestingly, although Proposition 2 holds true for any $ME > 0$ (given our assumptions), the incentives for asymmetrical efforts (overfocusing on the non-targeted) are not as strong as ME increases. Intuitively, although the targeted units already received training from the consulting partner (according to the level of e_c), both non-targeted and targeted units also benefit from the fixed effort e_f that the managers employ to oversee all units, which in turn is moderated by ME. As such, given the concavity of *f*, for higher levels of managerial engagement (i.e., $ME = ME^{\text{Higher}} > ME^{\text{High}} > 0$), the difference in $U(e_1, e_0)$ between employing optimal efforts $(e_0, e_1) = (e_0^*, e_1^*)$ or $e_0 = e_1$ is not as substantial[2]. Besides, if there are complementary effects between e_c and the e_1 (i.e., generating synergies between the consulting partner and manager efforts that would benefit only the targeted units), managers would have reduced incentives for employing asymmetrical efforts toward the non-targeted or would even focus on the targeted depending on the magnitude of those synergies. These alternative specifications point to boundary conditions on Proposition 2 that our data cannot test directly and that we leave for future work.

[2] More formally, the gains in utility (performance gap) deriving from optimal asymmetrical efforts are higher for positive but lower levels of ME:

$$\left(Y^{T=1} - Y^{T=0} \middle| e_c = \bar{e_c}, e_f = \bar{e_f}, ME = ME^{\text{High}} > 0, e_0^* > e_1^*\right) - \left(Y^{T=1} - Y^{T=0} \middle| e_c = \bar{e_c}, e_f = \bar{e_f}, ME = ME^{\text{High}} > 0, e_0 = e_1\right) > \left(Y^{T=1} - Y^{T=0} \middle| e_c = \bar{e_c}, e_f = \bar{e_f}, ME = ME^{\text{Higher}} > ME^{\text{High}} > 0, e_0^* > e_1^*\right) - \left(Y^{T=1} - Y^{T=0} \middle| e_c = \bar{e_c}, e_f = \bar{e_f}, ME = ME^{\text{Higher}} > ME^{\text{High}} > 0, e_0 = e_1\right).$$

Online Appendix B

Table B1 further details the main organizations and key individuals in our empirical setting, as introduced in the manuscript, summarizing their expected roles and interactions with other players within and outside their organizations under the collaborative agreement.

TABLE B1: Partnering organizations and key individuals

Organization and main actors	Expected role	Interactions with other players
Consulting partner: Unibanco Institute **Key individual actors:** Consultants	Unibanco Institute is a nonprofit organization known for its expertise in managerial practices in education. In this agreement, their consultants are expected to promote the implementation of such managerial practices by the recipient partner, which ultimately depend on the implementation by their multiple units (in this setting, schools)	Consultants interact directly with the State Departments of Education's middle managers (middle-level managers) and the targeted schools' management teams (principal, vice-principal, and pedagogical coordinator) (i.e., they do not interact directly with the non-targeted schools). Such interactions should enable consultants to promote the implementation of managerial practices, considering the scope of the original agreement (i.e., enhance the performance of the targeted units).
Recipient partners: State Departments of Education **Key individual actors:** Middle managers, who are middle-level managers working as civil servants	State Departments of Education are part of the state subnational governments in Brazil, responsible for the state public schools (in our setting, the states of Piauí and Espírito Santo). Each state is divided into several geographical jurisdictions, and a specific manager is responsible for overseeing the activities of the multiple schools under a given jurisdiction. Such managers, that we refer to as middle managers, are responsible for facilitating the implementation of practices in these schools, monitoring their performance.	In the agreement, middle managers are critical players who nurture relationships with all the other actors. These managers oversee the activities of both targeted and non-targeted schools within their jurisdiction. They hold leading positions and are at the middle level between schools and the head of the State Department of Education. They receive training from the consulting partner's consultants and, under the agreement, support and follow up on implementation in the targeted schools alongside the consultants; they have no authority to impose compliance (they do not control schools' staffing or budgets, and no sanctions were attached to the program). Their responsibility for the non-targeted schools comes from their regular position, and the agreement neither required nor prohibited applying the practices there: in non-targeted schools, whatever the school receives comes through the middle manager and the supervisory team.
Recipient partner organizational units: Targeted schools and Non-targeted schools (all public state schools) **Key individual actors:** School management team (principal, vice-principal, pedagogical coordinator), who received the frontline training	The public schools are the ultimate responsible for students' learning outcomes and represent the State Departments of Education organizational units. In our empirical setting, following a random sorting by middle managers' jurisdiction, schools are classified as 'targeted' or 'non-targeted.' While the management team of targeted schools is trained by the consulting partner and expected to implement the practices, non-targeted schools might have access to them only through their middle manager.	While the management teams of targeted schools interact both with the consultants working for the consulting partner (Unibanco Institute) and their department's middle-level manager (middle managers, State Department of Education), the non-targeted schools interact only with the middle manager. We note that interactions between schools are limited in our context (Teodorovicz et al., 2023).

Online Appendix C

Table C1 shows the instrument used to gauge the managerial engagement of middle managers (*Managerial Engagement*$_{i(m)}$). Interviews were conducted in May 2018, after the implementation period, by an external research team commissioned by the Unibanco Institute with the middle managers of 20 of the 26 school districts in the evaluation; all 11 in Espírito Santo and 9 of the 15 in Piauí (as further detailed in Table C2). The Institute commissioned the interviews because its evaluation had revealed substantial variation in outcomes across districts despite an identical training at the middle-manager level, and it sought to understand why. The interviews followed a structured protocol, and each dimension was scored with the rubrics reproduced in Table C1; the interviewers were not the consultants who had worked with the managers and did not have access to school- or district-level results. All items refer to our definition of managerial engagement as the extent to which managers craft the necessary routines and build the necessary working relationships needed to internalize and disseminate the practices introduced by the consulting partner across the schools under their responsibility.

The interview protocol comprises five dimensions that capture how the manager took up the introduced practices and the working relationships through which they were put into use. Those dimensions are: (i) understanding of the managerial role, from mere bureaucratic execution to active engagement with the schools, (ii) use of data-driven routines, from basic or bureaucratic planning to the systematic use of data to guide planning, (iii) nurturing close relationships with upper-level managers at the State Department of Education, (iv) nurturing close relationships with the supervisory staff (school supervisors) who follow up with schools, and (v) closely monitoring schools. While the first two capture the manager's own uptake of the practices (Foss & Mazzelli, 2025; Kor, 2003; Sadun et al., 2017), the latter three capture the working relationships through which the manager mobilized the individuals involved and disseminated the proposed practices (Briscoe et al., 2015; Caldwell et al., 2017). Altogether, such items are appropriate for measuring our central construct because they capture what the manager did with the practices during the implementation period.

The Cronbach's alpha of our final measure is 0.70 using the non-standardized covariance matrix and 0.75 when computed in standardized form (based on the correlation matrix). We report the standardized coefficient because, although all items range from 1 to 4 (for item 3, managers who

reported no meetings with the State Department receive the lowest score, as in the rubric), they rely on construct-specific rubrics (see Table C1 below) and exhibit different variances, making the correlation-based alpha more appropriate. The final index of managerial engagement is extracted from these five items using principal axis factoring at the manager level (n = 20), retaining one factor and computing regression scores; an alternative extraction by principal components (0.97) and a simple average of the standardized items (0.96) yield closely related scores, and the high/low and tercile groupings coincide for 16 of the 20 managers under either alternative (Costello & Osborne, 2019). A maximum-likelihood extraction was also attempted but produces a Heywood case—the uniqueness of the first item converges to zero, so the solution collapses onto that single item—which is a common outcome with 20 observations and one dominant item; we therefore do not use it as a comparison. The factor structure indicates a preponderant latent dimension: only the first factor has an eigenvalue above one (2.02), while the remaining factors present eigenvalues well below one (all ≤ 0.21), suggesting that they capture only limited residual variance rather than substantively distinct constructs. Table C3 reports the factor loadings and uniquenesses. The understanding of the managerial role loads most strongly (0.89), followed by data-driven routines (0.63), monitoring of schools (0.58), and relationships with supervisors (0.56), while the relationship with the State Department loads least (0.41). Because regression scoring weights each item by its loading relative to its uniqueness, and the first item has by far the lowest uniqueness (0.20), the index is dominated by the understanding of the managerial role (scoring coefficient 0.66, against 0.04–0.13 for the other four items) and correlates at 0.98 with that item alone. The index thus reflects, above all, how the manager understood and took up the role of promoting performance in the schools under his or her responsibility, with the data-driven routines and the working relationships entering at the margin. This is also why the simple average of the standardized items, which weights all dimensions equally, yields attenuated estimates (Online Appendix E, Table E7). The Kaiser–Meyer–Olkin measure of sampling adequacy is 0.64.

Table C1 presents the items considered in our final measure and the rubrics used to determine their final scores, along with average scores for each item across our full sample and split samples of targeted and non-targeted schools. Although this is measured at the manager level, we present averages per sample of targeted and non-targeted schools for the sake of transparency. With standard errors

clustered at the middle-manager level (the level at which the items vary), none of the differences between targeted and non-targeted schools is statistically significant. Descriptives for our final measure and other observables are presented per split sample in Online Appendix D.

TABLE C1: Analytic Grids Used to Measure Managerial Engagement

Index composition and score categories	Understanding of the Managers' Role	Implementation of Data-Driven Initiatives	Relationship with State Department of Education	Relationship with Supervisors	Monitoring Schools
Lowest score: 1 out of 4	Executes only documentation	No proper planning, mainly bureaucratic	None or few meetings	None or few meetings	None or few meetings
Score: 2 out of 4	Documents and presents some concerns regarding general pedagogical issues	Basic planning with no focus on students' learning	A meeting at least every 3 months	Meetings only when requested or less than monthly	A meeting at least every 3 months
Score: 3 out of 4	All previous concerns, and further mentions the importance of connections with schools	Planning focused on students' learning, but with no initiatives regarding data analyses	A meeting at least between 1 and 3 months	A monthly meeting at least is required	A meeting at least between 1 and 3 months
Highest score: 4 out of 4	All previous concerns, and emphasize the importance of close monitoring and feedback with schools	Planning focused on students' learning, implementing initiatives with data analyses	A monthly meeting at least	A meeting at least every two weeks is required	A monthly meeting at least
Average (full sample)	3.38	2.61	2.99	3.62	2.47
Average (targeted schools)	3.38	2.67	3.01	3.62	2.48
Average (non-targeted schools)	3.39	2.54	2.97	3.63	2.45

Table C2 provides descriptive information on the interviewed middle managers, documenting interview characteristics and their prior educational and administrative experience.

TABLE C2: Interview details and middle managers' characteristics

Manager ID	Interview duration	State Department	Managerial Engagement Index	Years of Experience in Education	Previous roles (Yes =1; No= 0)				
					Teacher	Instructional Coordinator	School Principal	Other Roles at the Regional Level	Other Roles at the State Department Level
1	0:56:13	Espírito Santo	0.89	30 to 39	1	0	0	1	1
2	0:35:48	Espírito Santo	-1.41	30 to 39	1	0	1	0	0
3	0:33:37	Espírito Santo	0.58	30 to 39	1	1	1	1	0
4	0:27:06	Espírito Santo	-1.64	10 to 19	0	0	0	0	0
5	0:47:10	Espírito Santo	0.77	20 to 29	1	1	0	1	0
6	1:33:09	Espírito Santo	0.82	30 to 39	1	1	1	1	0
7	0:38:12	Espírito Santo	-1.68	20 to 29	1	1	1	1	0
8	0:47:56	Espírito Santo	0.77	20 to 29	1	1	1	1	1
9	0:39:23	Espírito Santo	0.52	20 to 29	1	0	1	1	0
10	0:43:25	Espírito Santo	0.77	10 to 19	1	1	1	0	0
11	0:45:16	Espírito Santo	0.70	30 to 39	1	0	1	1	0
12	1:14:47	Piauí	0.69	20 to 29	1	1	0	1	0
13	0:47:55	Piauí	-0.18	10 to 19	1	1	1	1	1
14	0:56:13	Piauí	-0.27	20 to 29	1	1	1	1	0
15	2:04:00	Piauí	-0.74	30 to 39	1	1	0	0	0
16	0:37:40	Piauí	-1.44	10 to 19	1	0	0	1	0
17	1:01:15	Piauí	0.52	20 to 29	1	0	0	1	0
18	1:10:48	Piauí	-0.40	20 to 29	1	0	1	1	0
19	0:47:18	Piauí	0.52	30 to 39	1	0	1	1	0
20	1:02:14	Piauí	0.23	10 to 19	1	1	0	0	0

Notes: Descriptive information is based on structured interviews conducted in May 2018 by an external research team commissioned by the Unibanco Institute as part of the evaluation of the *Jovem de Futuro* program. These interviews covered the managerial characteristics reported above and provided the scores for the five engagement dimensions described in Table C1. Interviews averaged 54 minutes. All interviews were conducted between May 7 and May 14, 2018. We conducted mean-comparison t-tests between high and low managerial engagement, based on the binary measure described in the text. None of the previous professional roles show statistically significant differences across the two groups. Specifically, the p-values are as follows: Teacher ($p = 0.43$), Instructional Coordinator ($p = 0.16$), School Principal ($p = 0.86$), Other Roles at the Regional Level ($p = 0.32$), and Other Roles at the State Department Level ($p = 0.33$). These results indicate that prior role experience does not differ systematically between managers with high versus low managerial engagement.

TABLE C3: Factor loadings of the managerial engagement index

Dimension	Factor loading	Uniqueness
(1) Understanding of the managerial role	0.89	0.20
(2) Data-driven routines	0.63	0.60
(3) Relationship with the State Department	0.41	0.83
(4) Relationship with supervisors	0.56	0.69
(5) Monitoring of schools	0.58	0.66
Eigenvalue of the first factor (second)	2.02 (0.21)	

Notes: Principal axis factoring of the five items at the middle-manager level (n = 20), one retained factor, regression scoring. Cronbach's alpha 0.70 (raw) and 0.75 (standardized); Kaiser–Meyer–Olkin measure of sampling adequacy 0.64. Regression scoring coefficients: 0.66 (understanding of the managerial role), 0.12 (data-driven routines), 0.04 (relationship with the State Department), 0.13 (relationship with supervisors), 0.13 (monitoring schools). A principal-components extraction and a simple average of the standardized items yield scores that correlate at 0.97 and 0.96 with the index, respectively; a maximum-likelihood extraction produces a Heywood case and is not used.

Online Appendix D

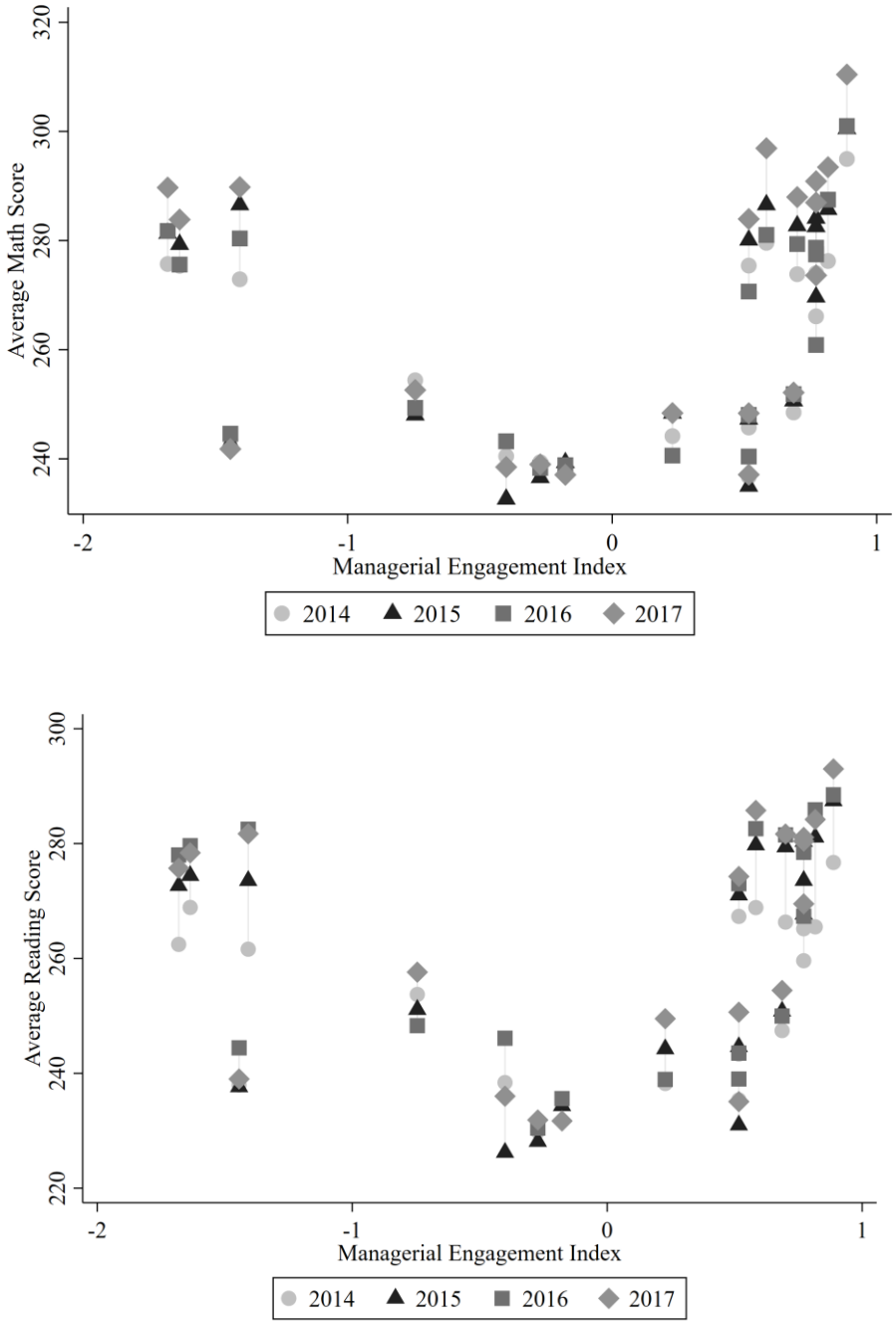
Table D1 reports means for targeted and non-targeted schools over the analysis period (2015–2017), with differences in means and two-sided p-values. Because the table pools post-treatment years, the higher Math and Reading scores of targeted schools reflect the effect of the intervention (H1) and the table is not a balance test. Targeted and non-targeted schools differ in some observable characteristics over this period (students' age and gender, computers per student, teachers with a college degree, and distance to the nearest school); all of these enter our specifications as controls, in first differences. There is no difference in managerial engagement between the two groups. Non-targeted schools are more isolated than targeted schools (more distant from other schools in general and from the closest targeted school), which should reinforce the role of managerial engagement in influencing the non-targeted schools.

TABLE D1: Difference in means for targeted and non-targeted schools

Variable	Mean (full sample)	Mean (targeted schools)	Mean (non-targeted schools)	Difference in means (p-values)	P-values considering clustered standard errors
Math Performance	265.96	268.93	261.52	7.41 (0.000)	(0.006)
Reading Performance	262.26	265.26	257.78	7.48 (0.000)	(0.003)
Managerial Engagement	-0.01	0.00	-0.01	0.01 (0.849)	(0.911)
Teachers per student (human resources)	0.10	0.10	0.11	-0.00 (0.233)	(0.390)
Teachers with college degree (human resources, alternative)	0.95	0.96	0.94	0.01 (0.004)	(0.055)
Computers per student (physical resources)	0.04	0.04	0.05	-0.00 (0.026)	(0.159)
Distance to closest targeted school	7.66	7.27	8.24	-0.97 (0.122)	(0.376)
Distance to the closest school	5.78	5.08	6.83	-1.74 (0.001)	(0.059)
Students' age	18.68	18.38	19.13	-0.75 (0.009)	(0.080)
Students' gender (male, %)	0.49	0.48	0.50	-0.02 (0.000)	(0.000)
Students' ethnicity (white, %)	0.21	0.21	0.21	0.00 (0.932)	(0.960)
Proportion of regular high school students	0.57	0.57	0.58	-0.01 (0.470)	(0.650)

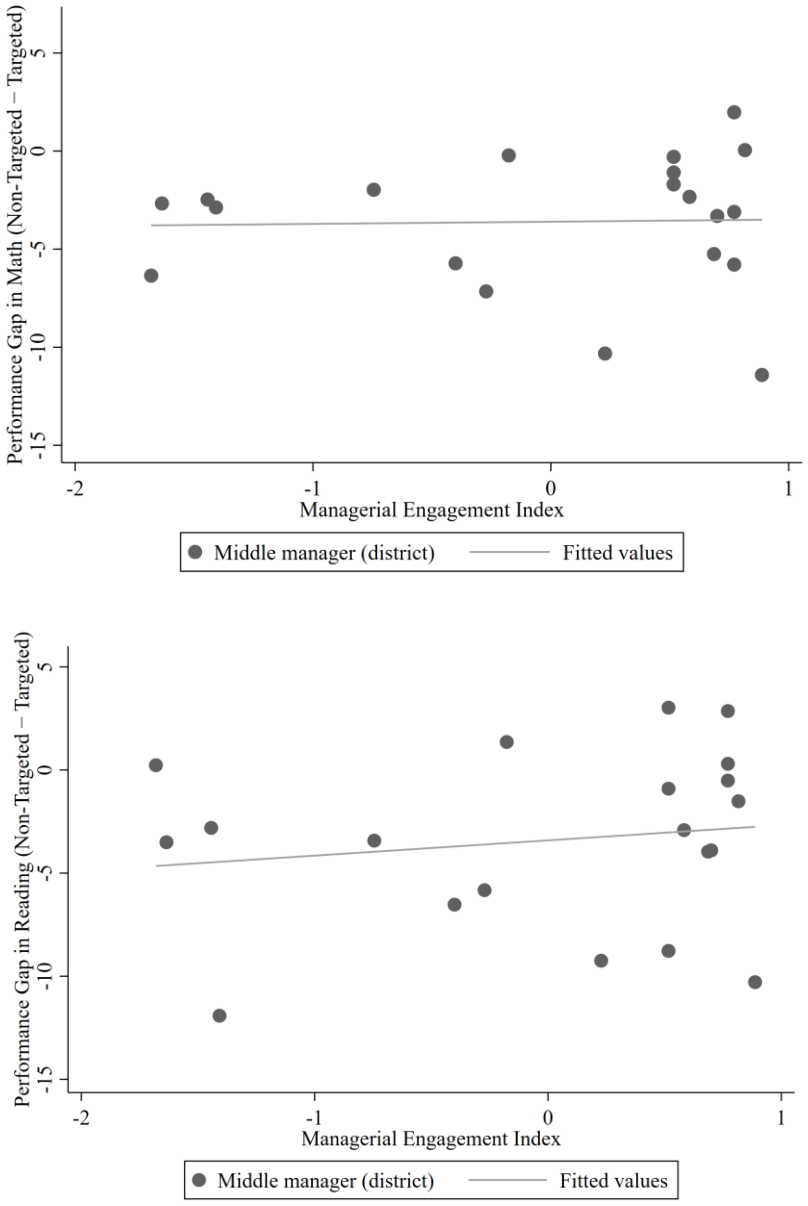
Notes: The table presents means for our full sample and split samples of targeted and non-targeted schools, with a difference in means according to t-tests and two-sided p-values. We additionally disclose p-values from regression-based tests with clustered standard errors, accounting for intra-cluster correlation, to provide more robust inferences when observations within schools should not be independent.

FIGURE D1: Test Score Levels by Managerial Engagement (Manager-Level Plots)



Notes: The figure displays average Math scores (top panel) and Reading scores (bottom panel) at the middle-manager level. Each point corresponds to the average test score of the schools supervised by a given middle manager in a given year. The horizontal axis reports the continuous managerial engagement index. Symbols differentiate years (2014–2017). Pre-treatment corresponds to 2014, while 2015–2017 represent post-treatment years. This allows for a direct visualization of within-manager changes over time as a function of managerial engagement.

FIGURE D2: Performance Gaps and Managerial Engagement



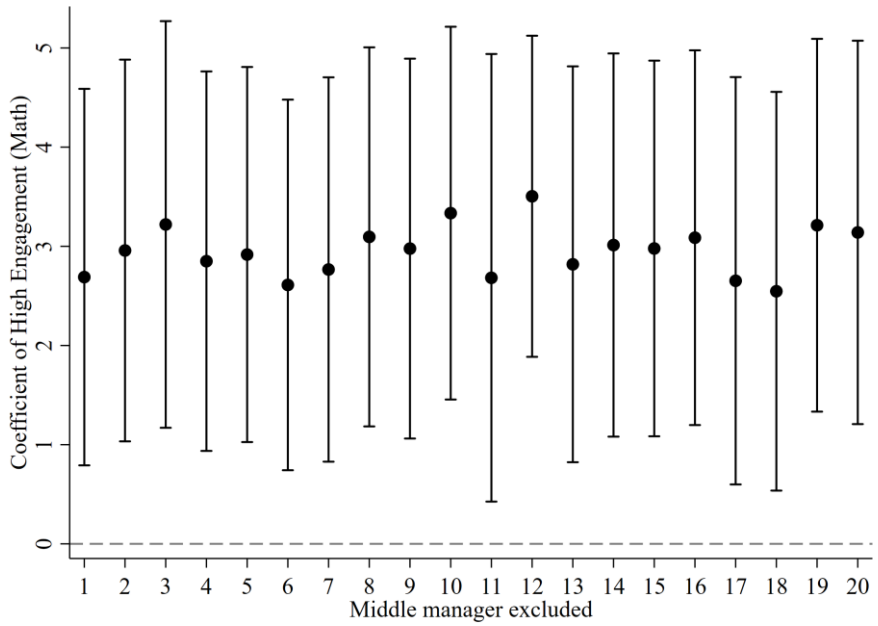
Notes: The figure displays manager-level performance gaps in Math (top panel) and Reading (bottom panel). Performance gaps are measured as the difference between non-targeted and targeted schools in the change from baseline performance (2014) to post-intervention performance (average of 2015–2017). The vertical axis reports these gaps, averaged across each middle manager's jurisdiction. Negative values indicate an advantage of targeted schools relative to non-targeted schools. The horizontal axis reports the continuous managerial engagement index. Each point represents one middle manager. The fitted line corresponds to a simple linear regression of the performance gap on managerial engagement.

Online Appendix E

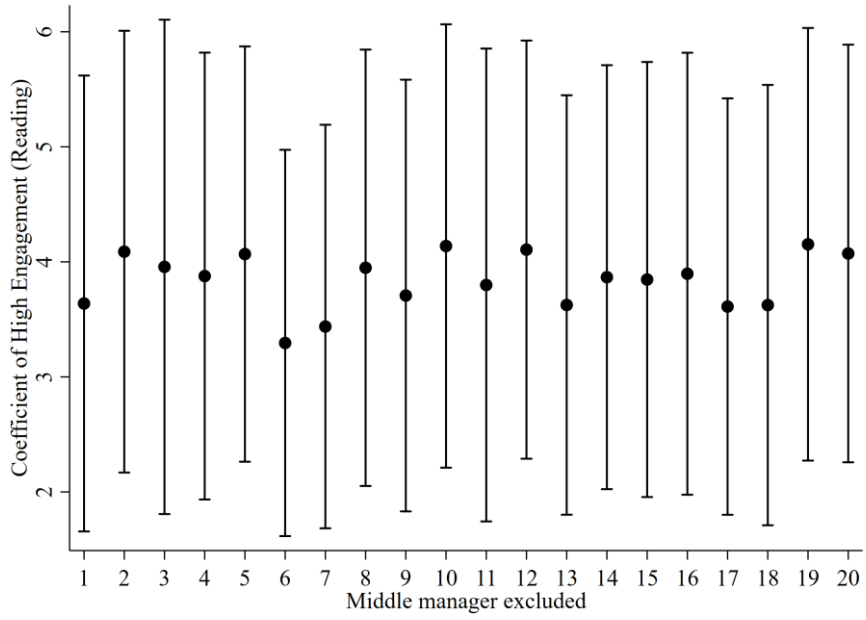
FIGURE E1

Leave-one-out Resampling Method: Dropping Middle Managers (School Districts)

a. Educational Outcome: Math



b. Educational Outcome: Reading



Notes: The figure shows plots of our point estimates (β2) on Math (panel a) and Reading (panel b) scores and corresponding 95% confidence intervals. Each point corresponds to the regression that drops all the schools of one middle manager; managers are numbered from 1 to 20 in the order of their district codes. These follow specification (3) as presented in Table 3 (considering a binary measure for Managerial Engagement—low vs. high).

TABLE E1: Resource Heterogeneity at the Recipient Partner Organizational Targeted Units

a. Human Capital Resource Heterogeneity

OLS clustered regressions		Math Performance			Reading Performance	
DV: ΔSchool performance	(5)	(6)	(7)	(5)	(6)	(7)
The role of human capital resources						
High Human Capital	-0.23	-0.68	0.66	-0.04	-0.36	-0.58
	(0.81)	(0.62)	(0.73)	(0.96)	(0.76)	(0.79)
Continuous measure for Managerial Engagement						
Managerial Engagement	0.82			0.59		
	(0.03)			(0.12)		
(H4) High Human C. * Managerial E.	-1.09			-0.52		
	(0.28)			(0.56)		
Binary measure: Low (Baseline) vs. High Managerial Engagement						
Managerial Engagement (High)		1.45			1.53	
		(0.17)			(0.13)	
(H4) High Human C. * Managerial E.(High)		1.36			1.11	
		(0.54)			(0.55)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High						
Managerial Engagement (Moderate)			0.53			-1.00
			(0.67)			(0.46)
Managerial Engagement (High)			2.08			1.09
			(0.09)			(0.33)
(H4) High Human C. * Managerial Engagement (Moderate)			-1.00			1.18
			(0.71)			(0.66)
(H4) High Human C. * Managerial Engagement (High)			-1.29			0.87
			(0.64)			(0.75)
Controls: Spillover between schools and additional controls						
Distance to closest targeted school	-0.07	-0.06	-0.07	-0.04	-0.03	-0.04
	(0.10)	(0.23)	(0.18)	(0.28)	(0.33)	(0.31)
Additional controls	Y	Y	Y	Y	Y	Y
Constant	4.32	3.38	3.23	4.58	3.62	4.32
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	672	672	672	672	672	672
R-squared	0.04	0.05	0.05	0.03	0.04	0.03

Notes: The table presents OLS regression estimates explaining school performance evolution regarding their student average Math and Reading scores. Table E1a mirrors Table 4a and investigates the human capital moderation effect (H4), but considering the subsample of targeted schools instead of the non-targeted. The control variables include all school, teachers, and student features, as described in our methodology section and Table 1 (additional controls are measured according to their first difference). P-values in parentheses (considering robust standard errors, clustered at the manager level).

b. Physical Resources Heterogeneity

OLS clustered regressions		Math Performance			Reading Performance	
DV: ΔSchool performance	(5)	(6)	(7)	(5)	(6)	(7)
The role of physical resources						
High Physical Resources	1.10	0.14	1.00	1.79	1.43	1.35
	(0.23)	(0.91)	(0.49)	(0.10)	(0.32)	(0.40)
Continuous measure for Managerial Engagement						
Managerial Engagement	0.46			0.51		
	(0.24)			(0.17)		
(H4) High Physical R. * Managerial E.	-0.20			-0.31		
	(0.78)			(0.66)		
Binary measure: Low (Baseline) vs. High Managerial Engagement						
Managerial Engagement (High)		1.12			1.68	
		(0.21)			(0.15)	
(H4) High Physical R. * Managerial E. (High)		1.83			0.58	
		(0.32)			(0.77)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High						
Managerial Engagement (Moderate)			0.33			-0.20
			(0.78)			(0.88)
Managerial Engagement (High)			1.30			0.90
			(0.15)			(0.18)
(H4) High Physical R. * Managerial Engagement (Moderate)			-0.32			-0.14
			(0.88)			(0.95)
(H4) High Physical R * Managerial Engagement (High)			0.23			0.91
			(0.91)			(0.61)
Controls: Spillover between schools and additional controls						
Distance to closest targeted school	-0.08	-0.08	-0.08	-0.06	-0.05	-0.06
	(0.10)	(0.16)	(0.13)	(0.17)	(0.15)	(0.12)
Additional controls	Y	Y	Y	Y	Y	Y
Constant	3.68	3.00	3.08	3.73	2.83	3.43
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Observations	672	672	672	672	672	672
R-squared	0.05	0.05	0.05	0.03	0.04	0.04

Notes: The table presents OLS regression estimates explaining school performance evolution regarding their student average Math and Reading scores. Table E1b mirrors Table 4b and investigates the physical resources moderation effect (H4), but considering the subsample of targeted schools instead of the non-targeted. The control variables include all school, teachers, and student features, as described in our methodology section and Table 1 (additional controls are measured according to their first difference). P-values in parentheses (considering robust standard errors, clustered at the manager level).

TABLE E2: School Performance in Levels (Post-Treatment Period Only)

OLS clustered regressions	Math Performance				Reading Performance			
DV: School performance	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Performance-enhancing managerial practices								
(H1) Originally Targeted	2.50	2.50	3.25	6.55	1.91	1.92	2.45	4.97
	(0.16)	(0.13)	(0.21)	(0.02)	(0.27)	(0.25)	(0.37)	(0.10)
Continuous measure for Managerial Engagement								
(H3) Originally Targeted * Managerial Engagement		-2.67				-1.30		
		(0.14)				(0.45)		
Binary measure: Low (Baseline) vs. High Managerial Engagement								
(H3) Originally Targeted * Managerial Engagement (High)			-1.73				-1.24	
			(0.60)				(0.69)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High								
(H3) Originally Targeted * Managerial Engagement (Moderate)				-5.23				-4.73
				(0.20)				(0.27)
(H3) Originally Targeted * Managerial Engagement (High)				-7.52				-4.78
				(0.04)				(0.18)
Controls: Spillover between schools and additional controls								
Distance to the closest targeted school	-0.04	-0.05	-0.04	-0.06	-0.09	-0.10	-0.09	-0.11
	(0.83)	(0.77)	(0.81)	(0.73)	(0.56)	(0.53)	(0.55)	(0.49)
Additional controls and middle managers FE	Y	Y	Y	Y	Y	Y	Y	Y
Constant	230.61	231.63	230.62	230.70	247.65	248.15	247.66	247.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	1,122	1,122	1,122	1,122	1,122	1,122	1,122	1,122
R-squared	0.67	0.67	0.67	0.67	0.64	0.64	0.64	0.64

Notes: The table presents OLS regression estimates of school performance measured by average student Math and Reading scores in levels, estimated using post-treatment observations only (2015–2017). By restricting the sample to the post-treatment period, the specification does not include interactions with a post indicator. The control variables include all school-, teacher-, and student-level characteristics described in the methodology section and Table 1. All models include middle manager fixed effects. P-values are reported in parentheses based on robust standard errors clustered at the manager level.

TABLE E3: School Performance in Levels: Difference-in-Differences Specification (2014–2017)

OLS clustered regressions	Math Performance		Reading Performance	
DV: School performance	(1)	(2)	(1)	(2)
Performance-enhancing managerial practices				
(H1) Originally Targeted * Post	3.30	3.27	3.06	3.04
	(0.00)	(0.00)	(0.01)	(0.01)
Originally Targeted	-0.87	-0.85	-1.12	-1.09
	(0.63)	(0.62)	(0.52)	(0.51)
Post	3.52	3.50	7.30	7.29
	(0.02)	(0.01)	(0.00)	(0.00)
Continuous measure for Managerial Engagement				
(H2) Managerial Engagement * Post		2.00		2.17
		(0.10)		(0.17)
(H3) Originally Targeted * Managerial E. * Post		-1.50		-1.15
		(0.07)		(0.18)
Originally Targeted * Managerial E.		-1.39		-0.32
		(0.49)		(0.85)
Distance to the closest targeted school	-0.00	-0.02	-0.07	-0.08
	(0.98)	(0.92)	(0.57)	(0.53)
Additional controls and middle managers FE	Y	Y	Y	Y
Constant	232.07	234.32	241.58	243.26
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	1,496	1,496	1,496	1,496
R-squared	0.65	0.65	0.61	0.61

Notes: The table presents OLS regression estimates of school performance measured by average student Math and Reading scores in levels, using the full sample period (2014–2017). The specification follows a standard difference-in-differences framework, using the post indicator. The control variables include all the same characteristics described in the methodology section and Table 1. All models include manager fixed effects. P-values are reported in parentheses based on robust standard errors clustered at the manager level.

TABLE E4: Asymmetrical Dissemination of Performance-Enhancing Managerial Practices (Alternative measure of 'performance' instead of 'distance' as control for spillover between schools)

OLS clustered regressions	Math Performance				Reading Performance			
DV: ΔSchool performance	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Performance-enhancing managerial practices								
(H1) Originally Targeted	1.27	1.28	1.45	2.18	1.03	1.05	1.54	2.18
	(0.06)	(0.04)	(0.05)	(0.01)	(0.10)	(0.09)	(0.06)	(0.06)
Continuous measure for Managerial Engagement								
(H2) Managerial Engagement		1.02				1.38		
		(0.11)				(0.05)		
(H3) Originally Targeted * Managerial Engagement		-0.61				-0.85		
		(0.21)				(0.21)		
Binary measure: Low (Baseline) vs. High Managerial Engagement								
(H2) Managerial Engagement (High)			2.63				3.72	
			(0.02)				(0.00)	
(H3) Originally Targeted * Managerial Engagement (High)			-0.60				-1.40	
			(0.52)				(0.11)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High								
(H2) Managerial Engagement (Moderate)				2.01				1.91
				(0.07)				(0.17)
(H2) Managerial Engagement (High)				2.86				3.79
				(0.05)				(0.01)
(H3) Originally Targeted * Managerial Engagement (Moderate)				-1.71				-1.97
				(0.11)				(0.08)
(H3) Originally Targeted * Managerial Engagement (High)				-1.34				-1.85
				(0.29)				(0.15)
Controls: Spillover between schools and additional controls								
ΔSchool performance (Math or Reading, according to the dependent variable) of	0.16	0.16	0.16	0.16	0.07	0.07	0.06	0.07
closest targeted school	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.06)	(0.05)
Additional controls	Y	Y	Y	Y	Y	Y	Y	Y
Constant	2.20	2.15	0.97	0.59	3.10	3.04	1.39	1.19
	(0.02)	(0.03)	(0.26)	(0.54)	(0.00)	(0.00)	(0.16)	(0.30)
Observations	1,108	1,108	1,108	1,108	1,108	1,108	1,108	1,108
R-squared	0.06	0.06	0.07	0.07	0.02	0.02	0.03	0.03

Notes: The table presents OLS regression estimates explaining school performance evolution regarding their student average Math and Reading. This table mirrors the specifications in Table 3, but instead of considering the distance to the closest targeted school, we consider the performance of the closest targeted school in the given year (either in Math or Reading, matching the dependent variable). The control variables include all school, teacher, and student features, as described in our methodology section and Table 1 (additional controls are measured according to their first difference). P-values in parentheses (considering robust standard errors, clustered at the manager level).

TABLE E5: Asymmetrical Dissemination of Performance-Enhancing Managerial Practices (adding interaction between targeted status and distance to the closest targeted school)

OLS clustered regressions	Math Performance				Reading Performance			
DV: ΔSchool performance	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Performance-enhancing managerial practices								
(H1) Originally Targeted	0.79	0.76	0.86	1.62	0.44	0.40	0.76	1.58
	(0.28)	(0.29)	(0.24)	(0.06)	(0.49)	(0.54)	(0.35)	(0.13)
Continuous measure for Managerial Engagement								
(H2) Managerial Engagement		0.99				1.34		
		(0.13)				(0.06)		
(H3) Originally Targeted * Managerial Engagement		-0.58				-0.85		
		(0.20)				(0.17)		
Binary measure: Low (Baseline) vs. High Managerial Engagement								
(H2) Managerial Engagement (High)			2.88				3.77	
			(0.01)				(0.00)	
(H3) Originally Targeted * Managerial Engagement (High)			-0.59				-1.23	
			(0.47)				(0.15)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High								
(H2) Managerial Engagement (Moderate)				2.03				1.99
				(0.07)				(0.15)
(H2) Managerial Engagement (High)				3.02				3.81
				(0.03)				(0.00)
(H3) Originally Targeted * Managerial Engagement (Moderate)				-1.90				-2.41
				(0.02)				(0.02)
(H3) Originally Targeted * Managerial Engagement (High)				-1.29				-1.81
				(0.26)				(0.12)
Controls: Spillover between schools and additional controls								
Distance to the closest targeted	-0.15	-0.16	-0.15	-0.17	-0.12	-0.12	-0.11	-0.13
	(0.01)	(0.01)	(0.02)	(0.01)	(0.13)	(0.09)	(0.10)	(0.06)
Distance to the closest targeted *	0.09	0.10	0.10	0.11	0.09	0.10	0.10	0.11
Originally Targeted	(0.08)	(0.05)	(0.03)	(0.04)	(0.14)	(0.10)	(0.07)	(0.07)
Additional controls	Y	Y	Y	Y	Y	Y	Y	Y
Constant	3.16	3.16	1.85	1.56	3.74	3.74	2.03	1.87
	(0.00)	(0.00)	(0.03)	(0.13)	(0.00)	(0.00)	(0.03)	(0.10)
Observations	1,122	1,122	1,122	1,122	1,122	1,122	1,122	1,122
R-squared	0.04	0.04	0.04	0.04	0.02	0.02	0.03	0.02

Notes: This table mirrors the specifications in Table 3, but instead of considering only the distance to the closest targeted school, we add the interaction term between the targeted status and distance to the closest targeted school. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1 (additional controls are measured according to their first difference). P-values in parentheses (considering robust standard errors, clustered at the manager level).

TABLE E6: Asymmetrical Dissemination of Performance-Enhancing Managerial Practices (Additionally controlling for middle managers' fixed effects)

OLS clustered regressions	Math Performance				Reading Performance			
DV: ΔSchool performance	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Performance-enhancing managerial practices								
(H1) Originally Targeted	1.28	1.28	1.42	1.72	1.05	1.05	1.57	1.73
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.02)	(0.05)
Continuous measure for Managerial Engagement								
(H3) Originally Targeted * Managerial Engagement		-0.27				-0.50		
		(0.59)				(0.36)		
Binary measure: Low (Baseline) vs. High Managerial Engagement								
(H3) Originally Targeted * Managerial Engagement (High)			-0.33				-1.22	
			(0.73)				(0.13)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High								
(H3) Originally Targeted * Managerial Engagement (Moderate)				-0.94				-1.12
				(0.42)				(0.32)
(H3) Originally Targeted * Managerial Engagement (High)				-0.38				-0.96
				(0.75)				(0.34)
Controls: Spillover between schools and additional controls								
Distance to closest targeted school	-0.08	-0.08	-0.08	-0.08	-0.06	-0.06	-0.06	-0.06
	(0.16)	(0.15)	(0.15)	(0.15)	(0.15)	(0.13)	(0.13)	(0.12)
Additional controls (including middle managers' fixed effects)	Y	Y	Y	Y	Y	Y	Y	Y
Constant	-0.20	-0.28	-0.26	-0.36	1.55	1.38	1.32	1.27
	(0.80)	(0.70)	(0.73)	(0.61)	(0.12)	(0.19)	(0.20)	(0.25)
Observations	1,122	1,122	1,122	1,122	1,122	1,122	1,122	1,122
R-squared	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04

Notes: The table presents OLS regression estimates explaining school performance evolution regarding their student average Math and Reading scores. This table mirrors the specifications in Table 3, but additionally considers middle managers' fixed effects as controls. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1 (additional controls are measured according to their first difference). P-values in parentheses (considering robust standard errors, clustered at the manager level).

TABLE E7: Asymmetrical Dissemination of Performance-Enhancing Managerial Practices (considering simple average for Managerial Engagement instead of factor)

OLS clustered regressions	Math Performance				Reading Performance			
DV: ΔSchool performance	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Performance-enhancing managerial practices								
(H1) Originally Targeted	1.49	1.48	1.98	1.72	1.13	1.14	1.74	1.77
	(0.03)	(0.02)	(0.01)	(0.05)	(0.08)	(0.08)	(0.05)	(0.12)
Continuous measure for Managerial Engagement								
(H2) Managerial Engagement		0.77				1.11		
		(0.31)				(0.13)		
(H3) Originally Targeted * Managerial Engagement		-0.38				-0.72		
		(0.57)				(0.36)		
Binary measure: Low (Baseline) vs. High Managerial Engagement								
(H2) Managerial Engagement (High)			1.82				2.42	
			(0.11)				(0.06)	
(H3) Originally Targeted * Managerial Engagement (High)			-1.13				-1.39	
			(0.22)				(0.14)	
Terciles of Managerial Engagement: Low (Baseline) vs. Moderate vs. High								
(H2) Managerial Engagement (Moderate)				0.28				0.09
				(0.85)				(0.96)
(H2) Managerial Engagement (High)				2.11				3.08
				(0.10)				(0.01)
(H3) Originally Targeted * Managerial Engagement (Moderate)				-0.15				-0.50
				(0.91)				(0.73)
(H3) Originally Targeted * Managerial Engagement (High)				-0.56				-1.42
				(0.62)				(0.26)
Controls: Spillover between schools and additional controls								
Distance to the closest targeted school	-0.08	-0.08	-0.07	-0.07	-0.04	-0.04	-0.03	-0.04
	(0.05)	(0.06)	(0.10)	(0.10)	(0.30)	(0.27)	(0.43)	(0.35)
Additional controls	Y	Y	Y	Y	Y	Y	Y	Y
Constant	2.76	2.73	1.87	1.85	3.33	3.31	2.15	2.14
	(0.01)	(0.01)	(0.12)	(0.08)	(0.00)	(0.00)	(0.09)	(0.04)
Observations	1,122	1,122	1,122	1,122	1,122	1,122	1,122	1,122
R-squared	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02

Notes: This table mirrors the specifications in Table 3, but instead of considering our main factor-based measure of managerial engagement, we rely on a z-scored and averaged index constructed from the five underlying dimensions. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1 (additional controls are measured according to their first difference). P-values in parentheses (considering robust standard errors, clustered at the manager level).