Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Headteacher experience, management quality and pupils' learning in Sub-Saharan African schools: evidence from PASEC 2019

Kouassi Sylvain Assienin

ksylvainassienin@gmail.com

Université Jean Lorougnon Guédé https://orcid.org/0009-0005-5704-4711

Christian K. N'da

Université Jean Lorougnon Guédé https://orcid.org/0000-0002-8642-2006

Auguste K. Kouakou

Université Jean Lorougnon Guédé https://orcid.org/0000-0002-4486-2046

Research Article

Keywords: school governance, learning outcomes, Sub-Saharan Africa, double machine learning, quantile regression, PASEC

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11586412/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Headteacher Experience, Management Quality and Learning in Sub-Saharan African Schools

Kouassi Sylvain Assienin

Abstract

School governance is widely promoted as a low-cost lever for learning in Sub-Saharan Africa, yet most evidence rests on in-sample prediction or on estimates that ignore the clustering of pupils in schools. Using the 2019 PASEC survey (25,850 pupils in 14 countries), we relate sixteen governance indicators to a continuous learning score in all schools, public schools and private schools. We combine OLS with HC3 errors, double machine learning (DML) with school-clustered inference and Holm's correction, decile quantile regression and a penalised-spline additive model (GAM), and we benchmark their predictive accuracy against XGBoost, Random Forest and Extra Trees on one common set of held-out schools. Three findings emerge. First, flexible models do not beat linear ones on unseen schools: in all schools OLS and median quantile regression reach a test R^2 of 30.6%, the GAM 29.2% and the tree-based models 24.4% to 26.2%; in private schools no model exceeds 7%. Second, once the sixteen tests of each sample are corrected, only the headteacher's years of teaching is robustly associated with learning in all schools (DML +34.3 points across its observed range) and, in public schools, together with management quality (+62.9 and +10.0 points, the latter at the 10% level only); nothing survives in private schools. Third, the association of teaching experience is positive at every decile of the score distribution and peaks near the median, and equality of slopes across deciles is rejected for it in all and public schools. The evidence concerns conditional associations in cross-sectional data, not causal effects.

Keywords: school governance; learning outcomes; Sub-Saharan Africa; double machine learning; quantile regression; PASEC

JEL codes: I21, I25, O15, C21

1. Introduction

Education is a pillar of human and economic development (Becker, 1964), and in Sub-Saharan Africa it is seen as a lever against poverty and inequality. Its returns, however, materialise only if schooling produces skills (Hanushek & Woessmann, 2023). The World Bank estimates that 89% of children in the region cannot read and understand a simple text at age ten, against 87% before the pandemic (World Bank Group, 2022). The challenge is no longer access alone but what Bold et al. (2017) call enrolment without learning.

The PASEC 2019 assessment (PASEC, 2020) measures this deficit at school level. It covered 14 French-speaking Sub-Saharan countries, testing more than 62,000 end-of-primary pupils and 21,000 early-primary pupils in over 3,500 schools. Only 48% of end-of-primary pupils reach the sufficient level in reading and 38% in mathematics. The average hides large gaps: the share of pupils below the reading threshold goes from 6.6% in Gabon to 82.5% in Madagascar, and in mathematics it exceeds 77% in five

countries. Côte d'Ivoire illustrates the paradox, with mean scores falling between 2014 and 2019 from 517.0 to 502.8 in French and from 475.7 to 454.0 in mathematics.

This diagnosis is not a shortage of funds. African countries spent on average 5% of GDP and 16% of public budgets on education between 2010 and 2017, above the recommended floors of 4% and 15% (African Development Bank Group, 2020). Spending and learning do not coincide: primary spending per child is 22% above the regional mean in Lesotho, where 97% of children are learning-poor, and 36% below it in Kenya, where the rate is 79% (World Bank Group, 2022). Niger, which made the largest PASEC gains, did not change average class size; it evaluated contract teachers, reformed teacher training and strengthened headteacher accountability. How resources are governed thus matters at least as much as their volume, which is the place of local governance in the education production function (Hanushek, 1979).

The school is the relevant level of analysis. In most PASEC 2019 countries more than half of the variance in reading scores lies between schools, and between-school inequality rose between 2014 and 2019 in all countries, without rural–urban, gender or wealth gaps explaining it (PASEC, 2020). Pupils in private schools score about 22 points higher at constant location, a gap that says nothing about mechanisms. The headteacher's work (recruiting, supervising, leading teaching) is central where only half of teachers reach the highest reading proficiency level and many pupils face hunger and work outside school. Theory points the same way: school management and the partnership between school, family and community shape learning (Epstein & Jansorn, 2004). We test this hypothesis rather than presume it, since the literature on parental involvement also documents heterogeneous effects.

PASEC reports are descriptive by construction and their factor analyses stay essentially bivariate. Econometric work on the joint role of leadership practices, staff management and stakeholder relations is scarce in this context, rarely uses methods able to capture non-linearities and interactions, and seldom separates the public and private sectors. A previous paper using the same data (Assienin et al., 2025) classified pupils with neural networks without a validation that separates schools and without inference on the indicators.

This paper asks which governance practices, and above all which headteacher characteristics, are robustly associated with learning once inference respects the school structure and the number of tests. We contribute in three ways. First, we estimate partial associations for sixteen governance indicators by double machine learning with school-grouped cross-fitting, compare them with OLS, and control the family-wise error rate with Holm's procedure. Second, we show with quantile regression and an additive model how associations vary along the score distribution and whether they are linear. Third, we compare six models (OLS, median quantile regression, the additive model, XGBoost, Random Forest and Extra Trees) on one common split that holds whole schools out of estimation, which gives a measure of what governance and controls explain that does not depend on school identity and shows whether flexible learners add anything to linear ones. The estimates are conditional associations, not causal effects.

In our data the headteacher's years of teaching is the only indicator robustly and positively associated with learning in all schools; in public schools it is joined by the quality of school management at the 10% level, and no governance indicator survives in private schools. The rest of the paper is organised as follows: Section 2 reviews the literature, Section 3 presents the models and data, Section 4 gives results and discussion, and Section 5 concludes with policy implications and limitations.

2. Literature review

2.1 Schools, headteachers and learning

The education production function treats learning as the output of school inputs (Hanushek, 1979). Later work showed that teachers matter more than class size (Rivkin et al., 2005) and that the gap between schools is largely a gap in how they operate. Using PASEC 2014 data from ten countries and a two-level model, Kadio (2023) finds that differences in learning are mostly explained by school characteristics, whose effects vary by school type and location. A follow-up on PASEC data from ten West and Central African countries finds that teacher experience and qualifications, class size and infrastructure matter, and that private schools and strong infrastructure are associated with better outcomes (Kadio, 2025). These studies establish that schools differ, but their regressors are mainly resources and teacher traits, not governance practices.

2.2 School management and leadership

Bloom et al. (2015) measured management practices in more than 1,800 schools in eight countries and found that higher management quality is strongly associated with better outcomes; about half of the gap between autonomous and regular government schools is explained by principal leadership and governance. A systematic review of causal evaluations in low- and middle-income countries finds a small but significant average effect of management interventions on learning, about 0.03 to 0.04 standard deviations, and identifies low principal take-up and a long causal chain from practice to learning as obstacles (Anand et al., 2023). The same review shows that principals in middle-income countries have less decision-making authority and less teaching experience than in high-income countries. On the leadership side, transformational leadership theory holds that a leader's credibility with teachers rests on professional experience (Leithwood et al., 2008), and human-resource management is treated as a condition of school results (Moisset et al., 2018). Decentralised management reforms show effects that depend on how far schools can act on their mandate (Bruns et al., 2011).

2.3 Pedagogical practices, recognition and relations

Self-determination theory predicts that recognising effort strengthens intrinsic motivation (Deci & Ryan, 2013), which motivates rewards for the best pupils. Teacher collaboration is linked to better results (Goddard et al., 2007), while the effect of parental involvement depends on its form: reactive contact following observed difficulties is associated with weaker results and preventive involvement with stronger ones (Hill & Tyson, 2009). Models of school–family–community partnerships (Epstein & Jansorn, 2004; Torre & Murphy, 2016) assume a favourable role for these links.

2.4 Method and the case for school-level validation

Machine learning is increasingly used in applied economics for prediction problems (Mullainathan & Spiess, 2017), and double machine learning allows inference on a low-dimensional parameter while nuisance functions are learned flexibly (Chernozhukov et al., 2018). Tree-based ensembles such as Random Forest (Breiman, 2001), Extra Trees (Geurts et al., 2006) and gradient boosting (Chen & Guestrin, 2016) are the usual flexible benchmarks; whether they improve on linear models for school outcomes is an empirical question. When data are hierarchical, cross-validation that ignores the grouping gives optimistic error estimates, and grouped folds are recommended (Roberts et al., 2017). When many

hypotheses are tested, familywise error control matters for credibility (List et al., 2019). Most PASEC studies use regression or multilevel models and do not combine these elements.

2.5 Gap

To our knowledge, no study of PASEC data evaluates flexible and linear models on held-out schools while testing a full set of governance indicators with clustered inference and multiple-testing correction. Our earlier paper (Assienin et al., 2025) applied neural networks to a binary performance outcome without such validation or inference. This paper fills that gap.

3. Methodology and data

3.1 Modelling

Let y_{isc} be the score of pupil i in school s of country c, g_s the vector of the sixteen governance indicators of school s, x_{isc} the controls and μ_c a country effect. Each pupil is observed once, and pupils of one school share g_s.

3.1.1 Validation on unseen schools. Write $y_{is} = f(x_{is}) + u_s + e_{is}$, where u_s is a school effect with variance σ_u^2 and e_{is} is independent noise with variance σ_e^2.

Proposition 1 (optimism of the pupil split). Suppose a flexible learner can recover u_s from the school-level covariates, which are constant within s. Under a random pupil split its expected test error is about σ_e^2; under a school split it is at least $\sigma_u^2 + \sigma_e^2$ minus the part of u_s explained by covariates.

Proof. Under the pupil split a test pupil's school appears in the training sample with probability close to one, so the learner can estimate $f(x) + u_s$ from the training pupils of that school, leaving the error e_{is}, of variance σ_e^2. Under the school split, the school of a test pupil is absent from training; u_s is independent of the training data, and the best predictor of u_s given the data has error variance σ_u^2 unless the covariates explain u_s. The test error is then $\sigma_u^2 + \sigma_e^2$ minus the part of u_s explained by covariates. ▫

The consequence for the design is that the evaluation must hold whole schools out. We draw one split in which 20% of schools (seed 12) form the test sample and the remaining 80% the training sample, and we use the same split for every model, so that differences in accuracy reflect the models and not the split. Min–max scaling constants are estimated on the training sample. Within it, hyperparameters are chosen by five-fold cross-validation with folds formed of whole schools, the same folds being used for every model. OLS, DML, quantile regression and the GAM are estimated, and their inference computed, on the training sample; the test sample is used once, to measure accuracy on schools never seen.

3.1.2 OLS with HC3 errors. The baseline is

$$y_{isc} = \alpha + g_s{}'\beta + x_{isc}{}'\gamma + \mu_c + \varepsilon_{isc}$$

estimated by least squares, $\hat{\beta} = (X'X)^{-1}X'y$. With heteroskedastic errors,

$$\mathrm{Var}(\hat{\beta}) = (X'X)^{-1}X' \,\mathrm{diag}(\sigma_i^2)\, X(X'X)^{-1}.$$

The HC3 estimator replaces σ_i^2 by $\hat{e}_i^2/(1-h_{ii})^2$, where h_{ii} is the leverage of observation i.

Justification. The residual satisfies $E\left[\hat{e}_i^2\right] = (1 - h_{ii})\sigma_i^2$ plus terms in the other σ_j^2 weighted by squared off-diagonal elements of the hat matrix. Dividing by $(1 - h_{ii})^2$ approximates the squared leave-one-out prediction error, which is the least biased in small samples among the HC family (MacKinnon & White, 1985). The standard errors are not clustered, which is why we rely on DML for school-clustered inference.

3.1.3 Double machine learning. For each governance indicator D, in turn, the partially linear model is

$$y = \theta D + g(X) + U, \quad E[U \mid D, X] = 0, \qquad D = m(X) + V, \quad E[V \mid X] = 0$$

where X contains all other covariates. Let $\ell(X) = E[y \mid X]$, so that $\ell(X) = \theta m(X) + g(X)$. The orthogonal score is

$$\psi(W; \theta, \eta) = \Big(y - \ell(X) - \theta\big(D - m(X)\big)\Big)\,\big(D - m(X)\big), \qquad \eta = (\ell, m)$$

Proposition 2 (Neyman orthogonality). At the true (θ_0, η_0), the derivative of $E[\psi]$ with respect to η in any direction $(\delta\ell, \delta m)$ is zero.

Proof. Differentiate ψ along $\eta_0 + r(\delta\ell, \delta m)$ at $r = 0$: $d\psi/dr = (-\delta\ell + \theta_0\,\delta m)\,V - U\,\delta m$, using $y - \ell - \theta_0 V = U$ and $D - m = V$. Because $E[V \mid X] = 0$ and $E[U \mid X] = 0$, both terms have zero conditional mean given X, since $\delta\ell$ and δm are functions of X, so the expectation is zero. ▫

Orthogonality makes $\hat{\theta}$ insensitive, to first order, to errors in the nuisance estimates, so they can be learned with machine learning (here LightGBM). With cross-fitting, $\hat{\eta}$ on each fold is estimated from the other folds; the estimator solves $\sum_i \psi\left(W_i; \hat{\theta}, \hat{\eta}\right) = 0$, i.e. $\hat{\theta} = \sum_i \hat{V}_i\,\hat{Y}_i / \sum_i \hat{V}_i^2$, with $\hat{Y} = y - \hat{\ell}(X)$ and $\hat{V} = D - \hat{m}(X)$. Folds consist of whole schools, five in number, so a school never helps predict its own pupils. The variance is the clustered sandwich

$$\hat{V}\left(\hat{\theta}\right) = \frac{\sum_s \left(\sum_{i \in s} \hat{\psi}_i\right)^2}{\left(\sum_i \hat{V}_i^2\right)^2}$$

which allows arbitrary correlation among pupils of one school. We read θ as a conditional association, not an effect: the model assumes no unobserved confounding.

3.1.4 Multiple testing. Let $p_{(1)} \le \cdots \le p_{(m)}$ be the ordered p-values of the $m = 16$ DML tests of a sample.

Proposition 3 (Holm, 1979). The procedure that rejects $H_{(1)}, \ldots, H_{(k-1)}$, where k is the smallest index with $p_{(k)} > \alpha/(m - k + 1)$, controls the probability of at least one false rejection at α.

Proof. Let T be the set of true nulls, with m_0 elements, and p^* the smallest p-value among them. If any true null is rejected, then $p^* \le \alpha/(m - k + 1)$ for the index k at which p^* sits, and since at most $m - m_0$ false nulls precede it, $k \le m - m_0 + 1$, so $p^* \le \alpha/m_0$. By Boole's inequality, $P(p^* \le \alpha/m_0) \le m_0 \cdot \alpha/m_0 = \alpha$. ▫

The adjusted p-value is $\tilde{p}_{(k)} = \max_{j \le k} \min\{1, (m - j + 1)\,p_{(j)}\}$. We call an association *robust* when it is significant in OLS and in DML and its Holm-adjusted p-value is below 10%. The 10% level reflects a family of sixteen tests with few independent schools; 5% results are reported alongside.

3.1.5 Quantile regression. For a quantile level τ, with check function $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$,

$$\hat{\beta}(\tau) = \underset{b}{\operatorname{argmin}} \sum_i \rho_\tau\,(y_i - x_i{}'b)$$

This estimator consistently estimates the conditional τ-quantile because the minimiser of $E[\rho_\tau(y-q)]$ satisfies $F(q) = \tau$ (first-order condition: $\tau - F(q) = 0$), and the linear specification applies this to $q = x'\beta(\tau)$ (Koenker & Bassett, 1978). A profile of $\hat{\beta}_j(\tau)$ that varies with τ shows that a variable changes the dispersion of the score and not only its level. Equality of a slope across quantiles is tested with a Wald statistic,

$$W = \left(R\hat{\beta}\right)'\left(R\hat{\Sigma}R'\right)^{-1}\left(R\hat{\beta}\right) \sim \chi^2(q)$$

under the null, where R holds q contrasts across quantile levels and $\hat{\Sigma}$ is the joint asymptotic covariance (Koenker–Bassett formula, Powell kernel, Hall–Sheather bandwidth). This covariance ignores correlation within schools, so p-values are optimistic. The fit of each quantile model is measured by the Koenker–Machado pseudo-R², $1 - \hat{V}(\tau)/\tilde{V}(\tau)$, the ratio of the minimised loss of the full model to that of an intercept-only model.

3.1.6 Additive model. The GAM replaces the linear terms of the five continuous variables by smooth functions f_j represented by cubic B-splines with K_j basis functions, $f_j(x) = \sum_k b_{jk}\ B_k(x)$, and estimated by penalised least squares,

$$\min_{\alpha,b} \sum_i (\,y_i - \alpha - \sum_j f_j\left(x_{ij}\right) - z_i{}'\delta)^2 + \lambda \sum_j \int\ (f_j{}''(t))^2 dt$$

where z_i collects the remaining (linear) covariates and λ sets the trade-off between fit and smoothness (Eilers & Marx, 1996); $\lambda \to \infty$ gives a straight line. We use the penalised generalised additive model of the statsmodels library (`GLMGam`) for estimation, prediction, partial-effect plots and inference, so that the model that is compared with the others is the model that is tested. The basis dimension is 10 (reduced for variables with few distinct values) and the smoothing weight is chosen from a logarithmic grid by school-grouped cross-validation. Wald tests with school-clustered standard errors are computed on the B-spline coefficients of each variable, so they test whether the variable is associated with the score, not whether the relation is non-linear.

3.1.7 Predictive benchmarks and their comparison. XGBoost (Chen & Guestrin, 2016), Random Forest (Breiman, 2001) and Extra Trees (Geurts et al., 2006) are fitted to the same training sample with the same covariates. Their hyperparameters are chosen on the common school-grouped folds from small documented grids: learning rate (0.03, 0.06), tree depth (3, 5) and number of trees (300, 600) for XGBoost; the share of variables tried at each split (0.33, 0.60) and the minimum leaf size (5, 20, 50) for Random Forest and Extra Trees. They are used for prediction only and carry no inference. To compare a model m with OLS on the test sample, we take the pupil-level difference of squared errors, $d_i = e^2_{i,\mathrm{OLS}} - e^2_{i,m}$, and test whether its mean (the "MSE gain") differs from zero with standard errors clustered by school, in the spirit of Diebold and Mariano (1995). A positive gain means that model m is more accurate than OLS.

3.1.8 Roles of the models. OLS, DML and Holm carry inference; quantile regression and the GAM describe heterogeneity and functional form; all six models are compared on predictive accuracy on held-out schools.

3.2 Data and coding

We use the 2019 PASEC survey (CONFEMEN), a grade-level assessment of reading and mathematics in French-speaking countries, with pupil, teacher and headteacher questionnaires. The database covers 25,850 pupils in 14 countries; 4,151 attend private schools and 21,073 public schools, the rest attending community schools, which are kept in the pooled sample only. The outcome is the average of the ten plausible values in reading and mathematics, on a scale of 0 to 1,000, kept continuous. Treating this average as an observed score ignores measurement uncertainty (Section 5.2). After the split, the training sample contains 20,599 pupils in 1,068 schools for all schools, 16,960 pupils in 858 schools for public schools and 3,386 pupils in 179 schools for private schools; the test samples contain 5,251 pupils in 268 schools, 4,113 pupils in 215 schools and 765 pupils in 45 schools.

The sixteen governance indicators come from the headteacher and teacher questionnaires and fall into three dimensions (Table 1). Controls are pupil age, teacher age, an indicator for a teacher without a diploma, school type (pooled sample only) and country fixed effects, with Benin as the reference. The five continuous variables (headteacher seniority, headteacher years of teaching, meetings organised by the headteacher, pupil age, teacher age) are rescaled to the interval 0 to 1 by min-max scaling, so a coefficient on them is the change in the score when moving from the minimum to the maximum observed value, not the effect of one more year. Binary indicators take the value 1 for the practice. Missing pupil and teacher ages coded 99 are replaced by the mean of the valid ages. All governance indicators are measured at school level.

Dimension	Indicator	Coding
School management	Quality of school management	Index, 0 to 2
	Good administrative management	Binary
	Human-resource management	Binary
	Management of the school environment	Binary
	Management of facilities and equipment	Binary
	Headteacher seniority	Years, rescaled 0–1
	Headteacher years of teaching	Years, rescaled 0–1
Pedagogical practices	Pedagogical management	Binary
	Frequency of pedagogical meetings	Binary
	Meetings organised by the headteacher	Count, rescaled 0–1
	Teacher inspection	Binary
	Rewards for the best pupils	Binary
Stakeholder relations	Relations among colleagues	Index, 0 to 2
	School–community relations	Index, 0 to 2
	School–parent relations	Index, 0 to 3
	School–pupil relations	Index, 0 to 3

Table 1. Governance indicators and coding. Source: PASEC 2019, author's classification.

4. Results and discussion

4.1 Descriptive statistics

The average score is 503 points (standard deviation 89) in the pooled sample, 494 in public and 557 in private schools (Table 2, Figure 1). Private-school pupils are younger (12.2 against 13.0 years), have more teachers without a diploma (30% against 4%) and are more often rewarded for good results (79% against 57%). Practices of pedagogical management (81%) and teacher inspection (97%) are almost universal, so they carry little variation; formal pedagogical meetings are rare (3%). Headteachers have on average 8 years of seniority and 9.4 years of teaching experience, with a wide dispersion (about 7 years). Country differences are large, from about 600 points in Gabon to about 430 in Chad.

Variable	All: mean (SD)	Public: mean (SD)	Private: mean (SD)
Quality of school management (0–2)	0.72 (0.65)	0.74 (0.64)	0.61 (0.66)
Good administrative management	0.66 (0.48)	0.66 (0.47)	0.65 (0.48)
Human-resource management	0.41 (0.49)	0.41 (0.49)	0.38 (0.49)
School environment management	0.39 (0.49)	0.38 (0.49)	0.44 (0.50)
Equipment management	0.37 (0.48)	0.36 (0.48)	0.47 (0.50)
Headteacher seniority (years)	8.11 (6.85)	8.20 (6.89)	7.72 (6.70)
Headteacher years of teaching (years)	9.37 (6.97)	9.43 (6.82)	9.26 (7.80)
Pedagogical management	0.81 (0.39)	0.82 (0.39)	0.79 (0.41)
Frequency of pedagogical meetings	0.03 (0.16)	0.02 (0.15)	0.05 (0.21)
Meetings organised by headteacher	3.10 (2.15)	3.19 (2.29)	2.60 (1.04)
Teacher inspection	0.97 (0.17)	0.97 (0.16)	0.97 (0.16)
Rewards for best pupils	0.60 (0.49)	0.57 (0.50)	0.79 (0.40)
Relations among colleagues (0–2)	1.36 (0.62)	1.37 (0.61)	1.27 (0.64)
School–community relations (0–2)	1.08 (0.67)	1.08 (0.67)	1.13 (0.67)
School–parent relations (0–3)	1.77 (0.75)	1.77 (0.75)	1.82 (0.75)
School–pupil relations (0–3)	2.02 (0.72)	2.00 (0.71)	2.20 (0.74)
Pupil age (years)	12.90 (1.66)	13.03 (1.57)	12.16 (1.80)
Teacher age (years)	39.11 (8.24)	39.40 (7.81)	37.87 (9.84)
Teacher without diploma	0.08 (0.27)	0.04 (0.19)	0.30 (0.46)
Score (0–1,000)	503.43 (88.56)	494.25 (84.70)	557.32 (88.07)
Pupils	25,850	21,073	4,151

Table 2. Descriptive statistics (unscaled values, full samples). Source: author's calculations, PASEC 2019.

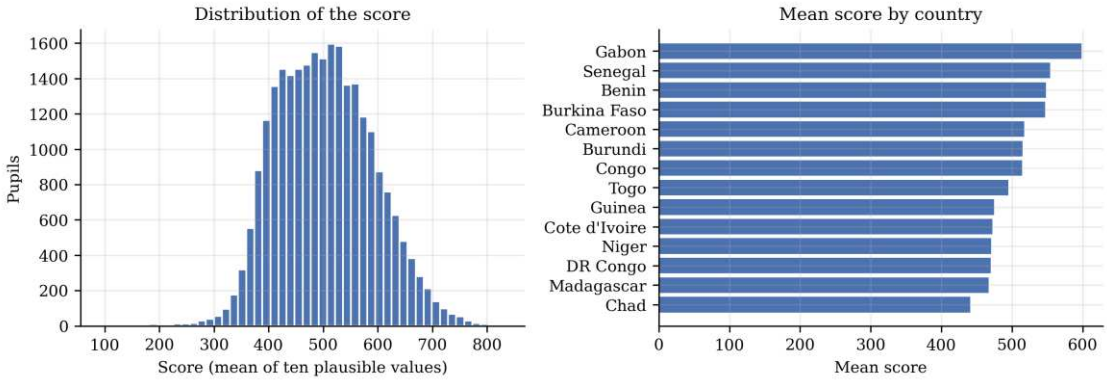


Figure 1. Distribution of the score and mean score by country, all schools. Source: author's calculations, PASEC 2019.

4.2 Performance of the models

Table 3 compares the six models on the same held-out schools. In all schools, OLS and median quantile regression reach a test R^2 of 30.6%, the GAM 29.2%, and the three tree-based models 25.4% (XGBoost), 24.4% (Random Forest) and 26.2% (Extra Trees). In public schools the ranking is the same: 26.1% for median quantile regression, 24.6% for OLS, 24.0% for the GAM and 19.7% to 21.8% for the tree-based models. In private schools no model explains more than 7% of the variance on unseen schools (6.8% for median quantile regression, 5.7% for OLS) and XGBoost does worse than the mean of the training sample (−0.1%). Figure 2 shows the same ranking for RMSE and MAE.

Sample	Model	CV R^2 (SD)	Test R^2	Test RMSE	Test MAE	MSE gain vs OLS (p)
All schools	OLS	22.7 (6.0)	30.6	72.5	57.3	—
	Quantile reg. (median)	21.6 (6.0)	30.6	72.5	56.8	+2 (0.978)
	GAM (splines)	23.6 (5.6)	29.2	73.2	57.9	−104 (0.012)
	XGBoost	23.7 (6.3)	25.4	75.2	59.7	−392 (0.001)
	Random Forest	24.2 (6.0)	24.4	75.7	60.2	−474 (0.001)
	Extra Trees	23.5 (6.6)	26.2	74.8	59.1	−332 (0.015)
Public schools	OLS	18.7 (5.2)	24.6	71.5	57.4	—
	Quantile reg. (median)	17.3 (5.6)	26.1	70.8	56.4	+98 (0.100)
	GAM (splines)	18.8 (5.5)	24.0	71.8	57.7	−45 (0.061)
	XGBoost	16.7 (4.8)	21.8	72.8	58.6	−190 (0.092)
	Random Forest	17.5 (5.1)	21.3	73.1	58.7	−225 (0.073)
	Extra Trees	18.8 (3.5)	19.7	73.8	59.4	−335 (0.013)
Private schools	OLS	3.9 (10.8)	5.7	81.6	65.5	—
	Quantile reg. (median)	3.9 (13.3)	6.8	81.1	65.1	+80 (0.662)
	GAM (splines)	4.2 (10.9)	6.1	81.4	65.4	+31 (0.326)
	XGBoost	−0.3 (5.9)	−0.1	84.0	68.0	−408 (0.582)
	Random Forest	5.2 (3.3)	5.3	81.8	66.6	−28 (0.961)
	Extra Trees	2.6 (6.6)	5.8	81.5	66.1	+10 (0.984)

Table 3. Predictive accuracy on schools held out of estimation (R^2 in %; RMSE and MAE in score points). CV R^2: mean and standard deviation over the five school-grouped folds of the training sample. MSE gain: mean over test pupils of the squared error of OLS minus that of the model, with the p-value of a test with school-clustered standard errors; a positive value favours the model. Source: author's calculations, PASEC 2019.

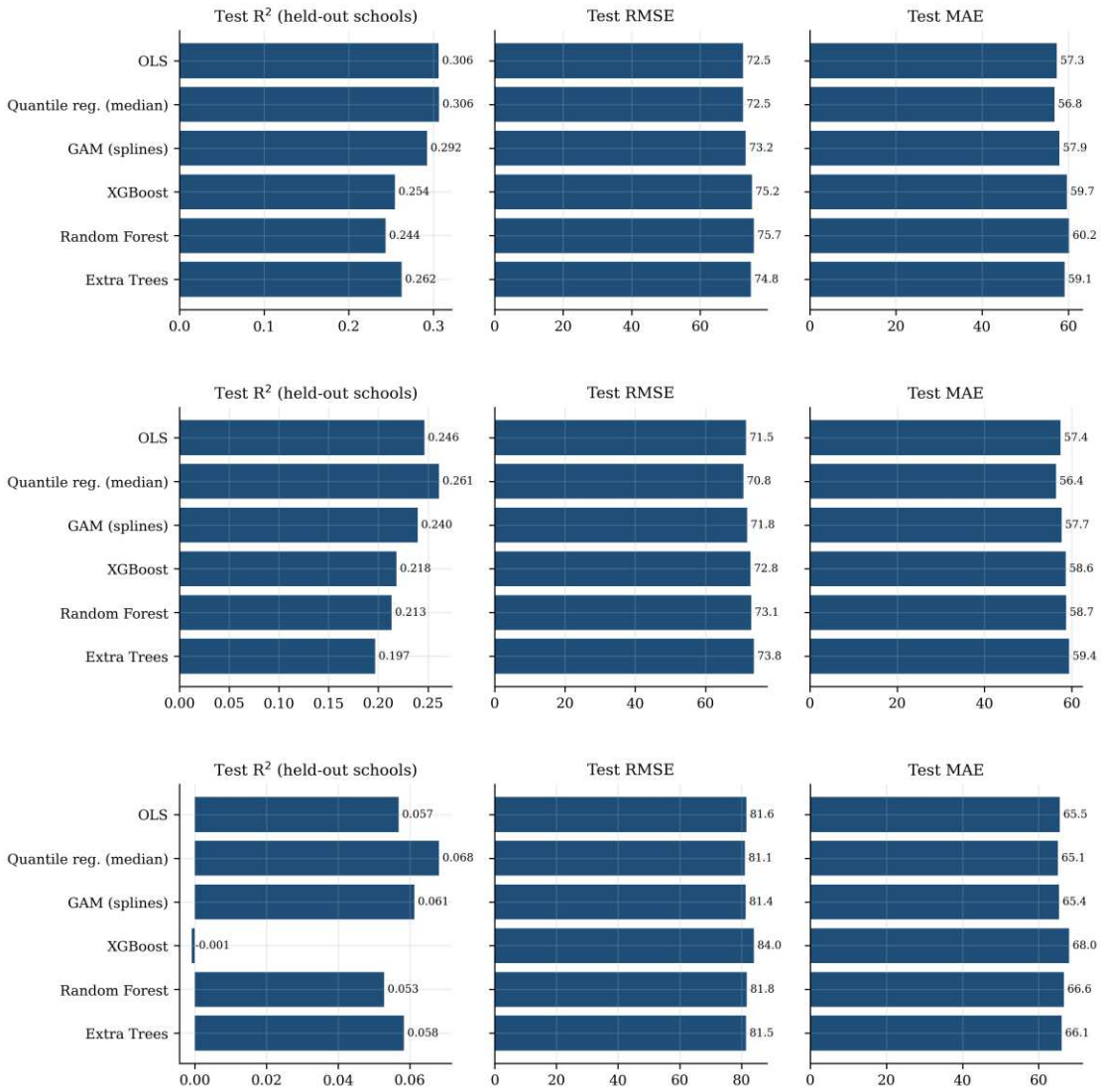


Figure 2. Test R², RMSE and MAE of the six models on held-out schools: all schools, public schools, private schools. Source: author's calculations, PASEC 2019.

The paired comparison with OLS gives a clear answer in all schools: the GAM, XGBoost, Random Forest and Extra Trees all have a significantly larger squared error than OLS ($p = 0.012$, 0.001, 0.001 and 0.015), while median quantile regression is indistinguishable from it ($p = 0.978$). In public schools the tree-based models and the GAM are also less accurate than OLS, with p-values between 0.061 and 0.092, except Extra Trees ($p = 0.013$), and median quantile regression is slightly more accurate ($p = 0.100$). In private schools, with only 45 test schools, no difference is significant, so the test cannot rank the models. On the training folds the picture is closer: the cross-validated R^2 of the flexible models (23.5% to 24.2% in all schools) is at most 1.5 points above that of OLS (22.7%), well within the fold-to-fold standard deviation of about 6 points, and the advantage does not carry over to schools held out in the test sample. Taken together, the results give no evidence that non-linearities or interactions among the observed covariates carry predictive information that a linear specification misses, so the linear models are an adequate basis for the inference that follows. The diagnostics of the OLS model point to heteroskedasticity (Breusch–

Pagan $p < 0.001$ in the three samples), which justifies robust standard errors, and to non-normal residuals in all and public schools (Jarque–Bera $p < 0.001$); the RESET test does not reject the linear specification in all schools ($p = 0.174$) or in private schools ($p = 0.181$) but does in public schools ($p = 0.029$), where the additive model nevertheless brings no gain. The largest variance inflation factor is 8.2 in all schools and below 3 in the other two samples.

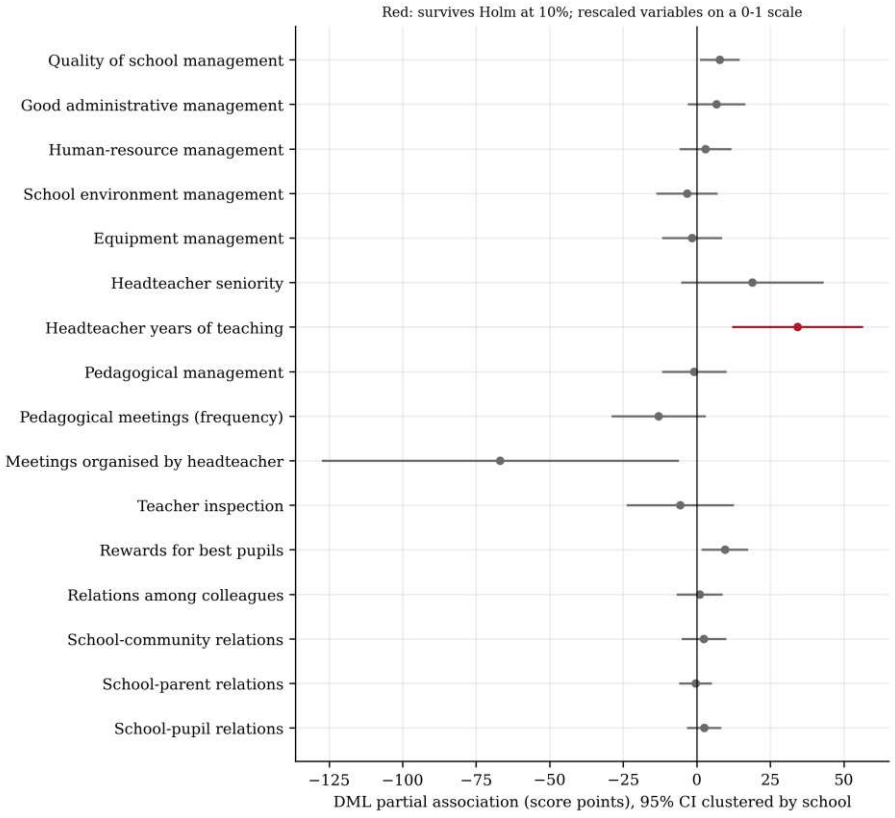
4.3 Results

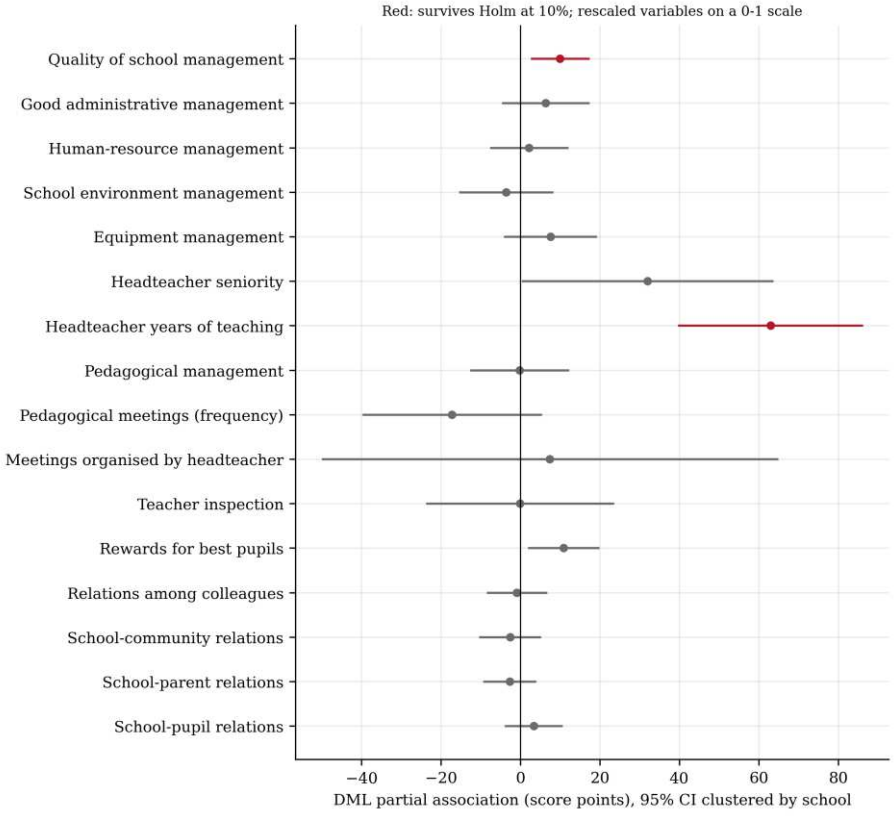
Associations of governance indicators. Table 4 reports the DML estimates on the training sample. Coefficients of the continuous variables are on the 0 to 1 scale, so they are changes across the observed range. Bold marks the associations that are robust, that is significant in OLS and in DML and with a Holm-adjusted p-value below 10%.

Indicator	All schools θ (Holm p)	Public θ (Holm p)	Private θ (Holm p)
Headteacher years of teaching (years)	**+34.3 (0.034)**	**+62.9 (<0.001)**	−7.8 (1.000)
Quality of school management (0–2)	+7.7 (0.243)	**+10.0 (0.091)**	−2.8 (1.000)
Headteacher seniority (years)	+18.8 (1.000)	+32.0 (0.595)	+20.0 (1.000)
Rewards for best pupils	+9.5 (0.203)	+10.9 (0.205)	+29.5 (0.307)
Meetings organised by headteacher	−66.9 (0.389)	+7.4 (1.000)	−52.6 (1.000)
Pedagogical management	−0.9 (1.000)	−0.3 (1.000)	+9.0 (1.000)
School–pupil relations (0–3)	+2.5 (1.000)	+3.3 (1.000)	+9.6 (1.000)
Human-resource management	+3.0 (1.000)	+2.2 (1.000)	−0.1 (1.000)
Good administrative management	+6.7 (1.000)	+6.3 (1.000)	−2.5 (1.000)
School environment management	−3.3 (1.000)	−3.6 (1.000)	+1.4 (1.000)
Equipment management	−1.6 (1.000)	+7.5 (1.000)	−15.1 (1.000)
Frequency of pedagogical meetings	−13.0 (1.000)	−17.2 (1.000)	+10.1 (1.000)
Teacher inspection	−5.6 (1.000)	−0.1 (1.000)	−23.9 (1.000)
Relations among colleagues (0–2)	+0.9 (1.000)	−0.9 (1.000)	+9.7 (1.000)
School–community relations (0–2)	+2.4 (1.000)	−2.6 (1.000)	+1.4 (1.000)
School–parent relations (0–3)	−0.5 (1.000)	−2.7 (1.000)	−7.6 (1.000)

Table 4. DML partial associations (score points) with Holm-adjusted p-values over the sixteen tests of each sample. Bold: robust (OLS and DML significant, Holm $p < 0.10$). Source: author's calculations, PASEC 2019.

Figure 3 plots the estimates with their 95% confidence intervals, red marking the associations that survive Holm's correction at 10%, and Table 5 gives the full DML output for all schools.


Red: survives Holm at 10%; rescaled variables on a 0-1 scale
Quality of school management
Good administrative management
Human-resource management
School environment management
Equipment management
Headteacher seniority
Headteacher years of teaching
Pedagogical management
Pedagogical meetings (frequency)
Meetings organised by headteacher
Teacher inspection
Rewards for best pupils
Relations among colleagues
School-community relations
School-parent relations
School-pupil relations
−125
−100
−75
−50
−25
0
25
50
DML partial association (score points), 95% CI clustered by school


Red: survives Holm at 10%; rescaled variables on a 0-1 scale
Quality of school management
Good administrative management
Human-resource management
School environment management
Equipment management
Headteacher seniority
Headteacher years of teaching
Pedagogical management
Pedagogical meetings (frequency)
Meetings organised by headteacher
Teacher inspection
Rewards for best pupils
Relations among colleagues
School-community relations
School-parent relations
School-pupil relations
−40
−20
0
20
40
60
80
DML partial association (score points), 95% CI clustered by school

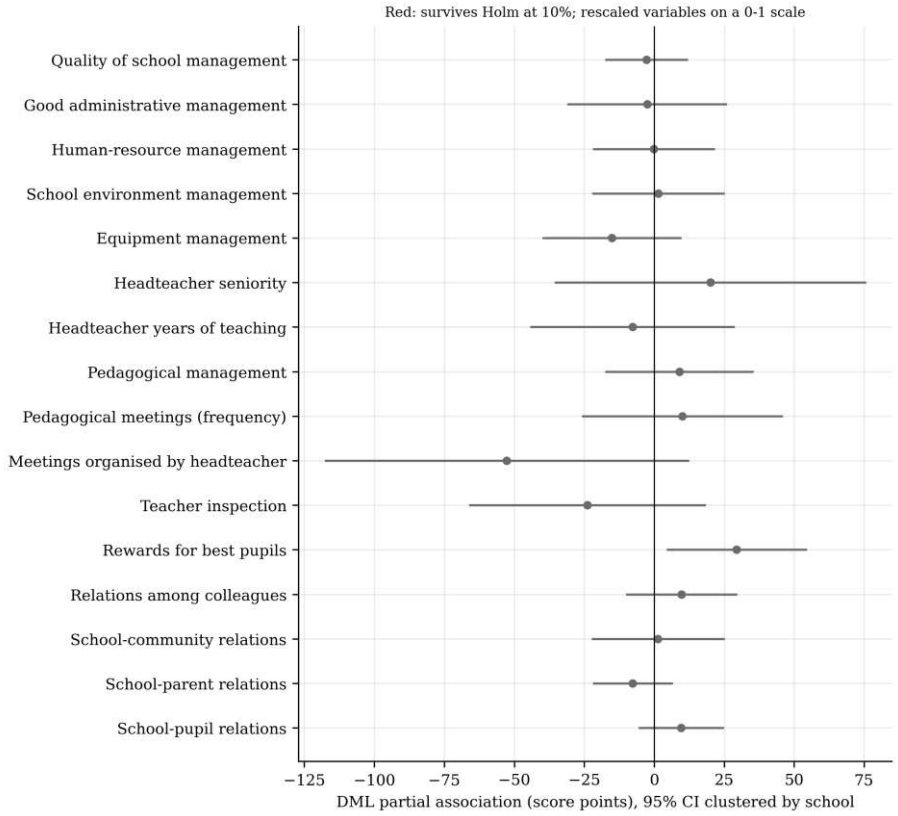


Figure 3. DML partial associations and 95% confidence intervals (school-clustered): all schools, public schools, private schools. Source: author's calculations, PASEC 2019.

Indicator	θ	SE	95% CI	p	Holm p	Holm 5%	Holm 10%
Quality of school management	+7.7	3.3	[1.4, 14.1]	0.017	0.243	no	no
Good administrative management	+6.7	4.8	[−2.7, 16.0]	0.164	1.000	no	no
Human-resource management	+3.0	4.3	[−5.5, 11.4]	0.490	1.000	no	no
School environment management	−3.3	5.1	[−13.4, 6.7]	0.516	1.000	no	no
Equipment management	−1.6	5.0	[−11.5, 8.2]	0.745	1.000	no	no
Headteacher seniority	+18.8	12.1	[−5.0, 42.6]	0.121	1.000	no	no
Headteacher years of teaching	+34.3	11.2	[12.4, 56.2]	0.002	0.034	yes	yes
Pedagogical management	−0.9	5.4	[−11.5, 9.7]	0.869	1.000	no	no
Pedagogical meetings (frequency)	−13.0	8.0	[−28.6, 2.6]	0.104	1.000	no	no
Meetings organised by headteacher	−66.9	30.8	[−127.2, −6.5]	0.030	0.389	no	no
Teacher inspection	−5.6	9.1	[−23.5, 12.2]	0.537	1.000	no	no
Rewards for best pupils	+9.5	3.9	[2.0, 17.1]	0.014	0.203	no	no
Relations among colleagues	+0.9	3.8	[−6.5, 8.4]	0.805	1.000	no	no
School–community relations	+2.4	3.7	[−4.9, 9.6]	0.523	1.000	no	no
School–parent relations	−0.5	2.6	[−5.6, 4.7]	0.864	1.000	no	no
School–pupil relations	+2.5	2.8	[−3.0, 8.0]	0.381	1.000	no	no

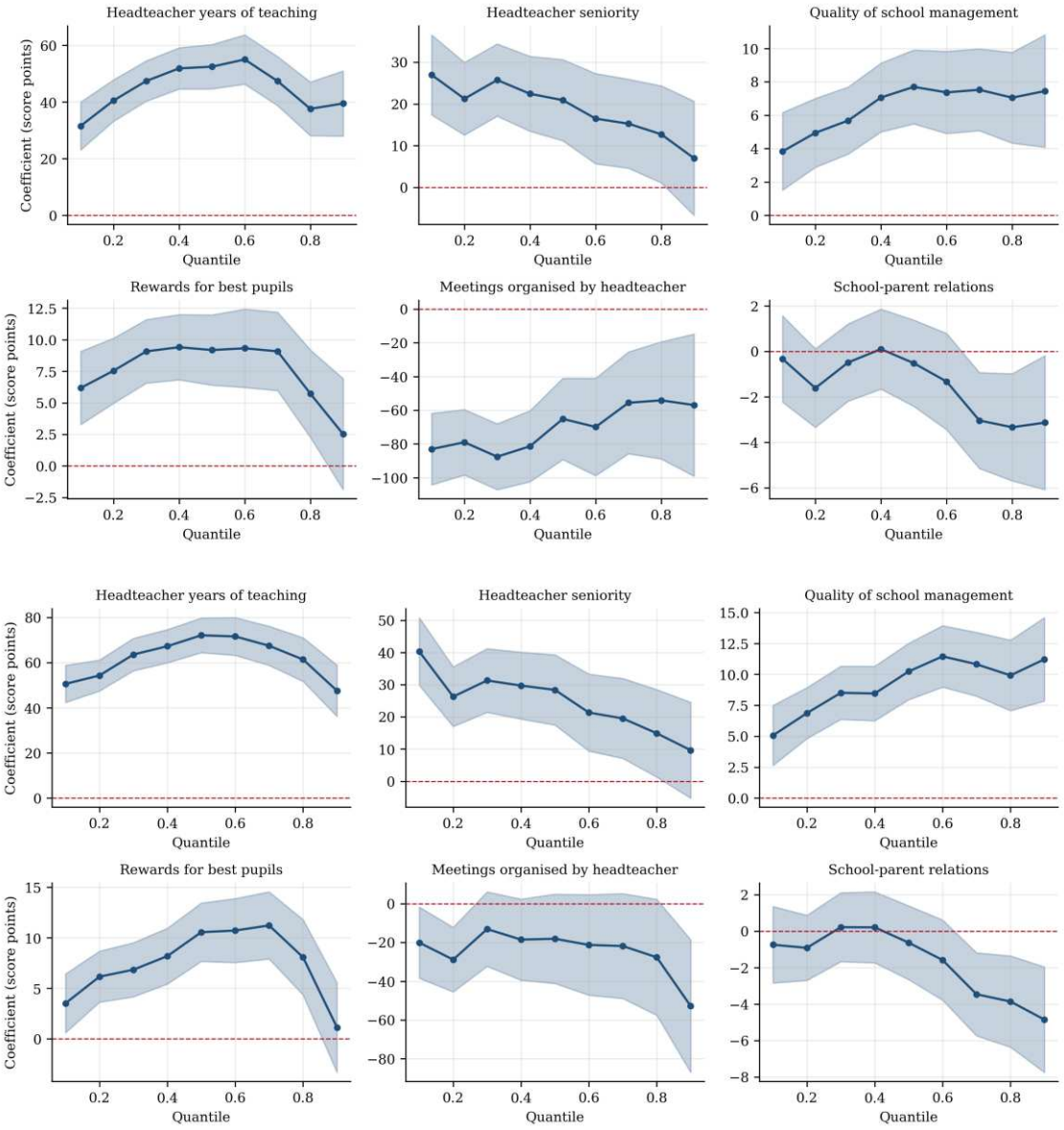
Table 5. DML estimates for all schools: standard errors, 95% confidence intervals and Holm-adjusted p-values. Source: author's calculations, PASEC 2019.

In all schools, the headteacher's years of teaching is the only robust association: +34.3 points across the observed range (Holm p = 0.034), which survives at the 5% level. Taking the observed range as 38 years,

one standard deviation (about 7 years) corresponds to about 6 points, or 0.07 standard deviations of the score. Management quality (+7.7), rewards for the best pupils (+9.5) and meetings organised by the headteacher (−66.9) are significant before correction ($p = 0.017$, 0.014 and 0.030) but their adjusted p-values are 0.24, 0.20 and 0.39. In public schools, years of teaching is again robust and larger (+62.9; Holm $p < 0.001$) and management quality is robust at the 10% level only (+10.0; Holm $p = 0.091$); headteacher seniority (+32.0, $p = 0.046$) and rewards (+10.9, $p = 0.015$) do not survive (adjusted $p = 0.595$ and 0.205). The pattern is informative: once years of teaching is controlled for, seniority in the post of headteacher is not robustly associated with learning in any sample, which suggests that classroom experience, rather than time spent as a headteacher, is what matters. The two variables are measured on the same scale but are correlated, so the comparison between them should be read with the width of the intervals in mind. The estimate for years of teaching is a range effect and should not be compared point for point with that of management quality, whose scale is 0 to 2.

In private schools nothing survives. Rewards for the best pupils has the largest estimate (+29.5, $p = 0.019$) but an adjusted p-value of 0.31. With 179 training schools, the clustered intervals are wide (several tens of points for the continuous variables), so the absence of robust results reflects lack of precision rather than evidence of no association. The role of clustering is large: in all schools OLS with HC3 errors flags 13 of the 16 indicators as significant at 10%, against only five when standard errors are clustered by school, and the DML-based Holm procedure retains one.

Heterogeneity along the distribution. Figure 4 gives the decile profiles with their confidence bands for the six most important variables of each model, and Table 6 the coefficients of the governance variables with robust or near-robust DML results. In all schools the association of the headteacher's years of teaching is positive at every decile, rising from +31.6 points at the first decile to +55.0 at the sixth and falling to +37.6 at the eighth; in public schools it is positive throughout, from +50.6 at the first decile to a maximum of +72.1 at the median and +47.6 at the ninth. In private schools the coefficients are negative at every decile (between −26.9 and −5.4), but they are not significant in DML and their equality across deciles is not rejected. The association of seniority decreases from +27.0 to +7.0 across deciles in all schools and from +40.3 to +9.6 in public schools, so the larger the score, the weaker the association. The global Wald test rejects equal slopes across the nine deciles in every sample ($p < 0.0001$), but mostly because of country and pupil-age indicators. For governance variables (Table 7), equality is rejected after Holm adjustment for the headteacher's years of teaching in all schools ($\chi^2(8) = 34.6$; adjusted $p = 0.001$) and in public schools (35.2; 0.001), and in the same two samples for school environment management (39.8 and 40.8; both below 0.001) and teacher inspection (24.7, 0.044; and 26.3, 0.022); in public schools also for rewards (34.2; 0.001), whose profile is an inverted U (+3.5 at the first decile, +11.2 at the seventh, +1.1 at the ninth). In private schools only pedagogical management rejects equality (34.7; 0.001). School environment management has a coefficient that becomes more negative along the distribution in all schools (from −3.3 to −20.4) and is significant in OLS, but its DML estimate is small and insignificant (−3.3, $p = 0.516$), so we do not read it as a robust association. The Koenker–Machado pseudo-R^2 ranges from 0.134 to 0.179 across deciles in all schools (highest at the fourth decile), from 0.101 to 0.168 in public schools and from 0.146 to 0.188 in private schools, where it is highest in the tails. Since the Wald covariance ignores clustering, these tests are indicative.


Headteacher years of teaching
Headteacher seniority
Quality of school management
Rewards for best pupils
Meetings organised by headteacher
School-parent relations
Coefficient (score points)
Quantile

Headteacher years of teaching
Headteacher seniority
Quality of school management
Rewards for best pupils
Meetings organised by headteacher
School-parent relations
Coefficient (score points)
Quantile

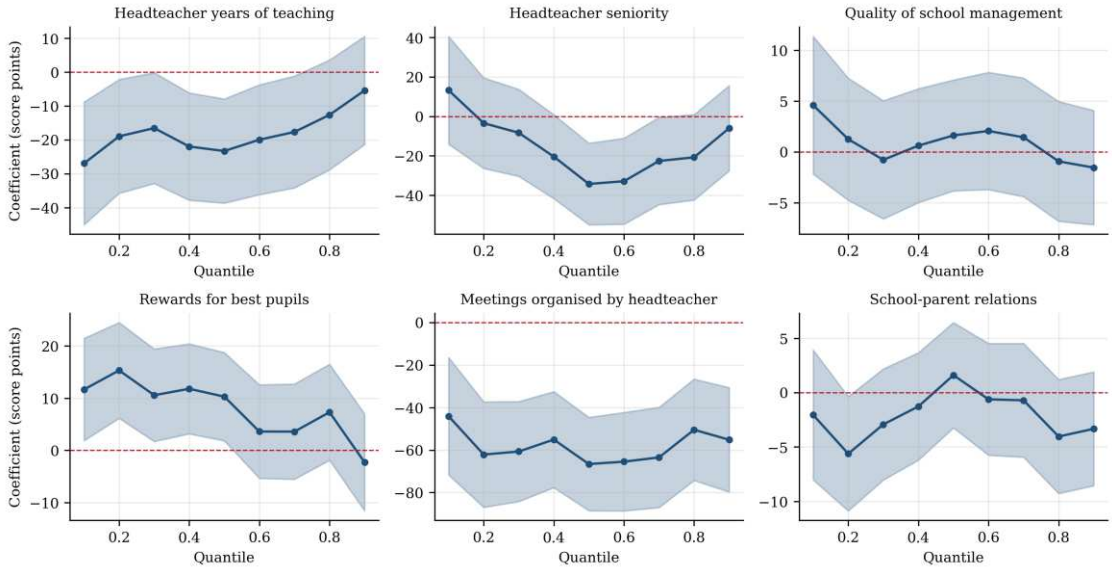


Figure 4. Quantile-regression coefficients by decile with confidence bands, six most important variables of each model: all schools, public schools, private schools. Source: author's calculations, PASEC 2019.

Variable	τ=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Headteacher years of teaching (All)	+31.6	+40.5	+47.4	+51.9	+52.5	+55.0	+47.4	+37.6	+39.5
Quality of school management (All)	+3.8	+4.9	+5.7	+7.1	+7.7	+7.4	+7.5	+7.0	+7.4
Rewards for best pupils (All)	+6.2	+7.5	+9.1	+9.4	+9.2	+9.3	+9.1	+5.7	+2.5
Headteacher seniority (All)	+27.0	+21.2	+25.7	+22.4	+20.9	+16.5	+15.3	+12.7	+7.0
Headteacher years of teaching (Public)	+50.6	+54.3	+63.6	+67.3	+72.1	+71.6	+67.5	+61.4	+47.6
Quality of school management (Public)	+5.1	+6.9	+8.5	+8.5	+10.2	+11.4	+10.8	+9.9	+11.2
Rewards for best pupils (Public)	+3.5	+6.2	+6.8	+8.2	+10.6	+10.7	+11.2	+8.1	+1.1
Headteacher seniority (Public)	+40.3	+26.3	+31.3	+29.7	+28.4	+21.3	+19.5	+14.9	+9.6
Headteacher years of teaching (Private)	−26.9	−19.0	−16.5	−22.0	−23.3	−19.9	−17.7	−12.6	−5.4
Rewards for best pupils (Private)	+11.7	+15.3	+10.6	+11.8	+10.3	+3.6	+3.6	+7.3	−2.3

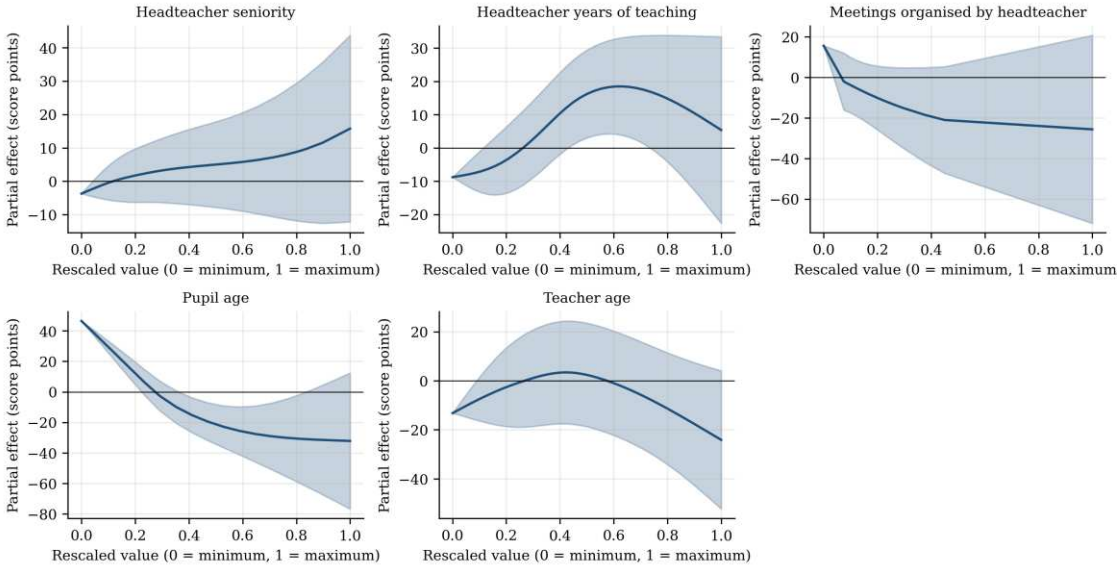
Table 6. Quantile-regression coefficients by decile (score points; rescaled variables on the 0–1 scale). Source: author's calculations, PASEC 2019.

Indicator	All: $\chi^2(8)$ (Holm p)	Public: $\chi^2(8)$ (Holm p)	Private: $\chi^2(8)$ (Holm p)
Quality of school management	10.0 (1.000)	22.7 (0.065)	3.9 (1.000)
Good administrative management	6.5 (1.000)	13.4 (0.790)	11.4 (1.000)
Human-resource management	3.3 (1.000)	9.1 (1.000)	9.5 (1.000)
School environment management	39.8 (<0.001)	40.8 (<0.001)	7.6 (1.000)
Equipment management	19.5 (0.258)	19.9 (0.141)	3.0 (1.000)
Headteacher seniority	6.7 (1.000)	18.0 (0.255)	17.4 (0.717)
Headteacher years of teaching	34.6 (0.001)	35.2 (0.001)	3.7 (1.000)
Pedagogical management	7.4 (1.000)	12.4 (0.936)	34.7 (0.001)
Pedagogical meetings (frequency)	16.7 (0.639)	10.7 (1.000)	5.5 (1.000)
Meetings organised by headteacher	6.1 (1.000)	11.2 (1.000)	4.3 (1.000)
Teacher inspection	24.7 (0.044)	26.3 (0.022)	6.4 (1.000)
Rewards for best pupils	15.0 (0.882)	34.2 (0.001)	8.4 (1.000)
Relations among colleagues	21.4 (0.134)	23.2 (0.055)	11.8 (1.000)
School–community relations	6.5 (1.000)	10.4 (1.000)	8.9 (1.000)

Indicator	All: $\chi^2(8)$ (Holm p)	Public: $\chi^2(8)$ (Holm p)	Private: $\chi^2(8)$ (Holm p)
School–parent relations	15.6 (0.832)	16.1 (0.456)	9.1 (1.000)
School–pupil relations	14.2 (1.000)	15.6 (0.479)	13.2 (1.000)

Table 7. Wald tests of equal slopes across the nine deciles, with Holm-adjusted p-values over the variables of each model. Source: author's calculations, PASEC 2019.

Functional form. The GAM adds little to OLS. In all schools the smoothing weight selected by school-grouped cross-validation (1.0) lowers the cross-validated RMSE by only 0.23 points relative to the straight-line limit (77.54 against 77.78), and in public schools the gain is 0.08 points (weight 3.2); in private schools the criterion keeps falling up to the end of the grid, so the GAM is a linear model there. On held-out schools the GAM is less accurate than OLS in all schools (test R^2 29.2% against 30.6%; p = 0.012). The partial effects (Figure 5) show a steep decline of the score with pupil age at young ages, flattening afterwards in all schools, and a rise then a plateau or a decline for the headteacher's years of teaching, with wide bands; they do not suggest pronounced non-linearities in the governance variables. Wald tests with clustered errors (Table 8) confirm that the headteacher's years of teaching is associated with the score in all schools (p = 0.006) and in public schools (p < 0.001), not in private schools (p = 0.422); pupil age is associated with the score in the three samples (p < 0.001). Meetings organised by the headteacher are only marginal in all schools (p = 0.099). Without clustering, the p-values of all five continuous variables are below 0.001 in all and public schools, which shows how much inference ignoring schools overstates precision.



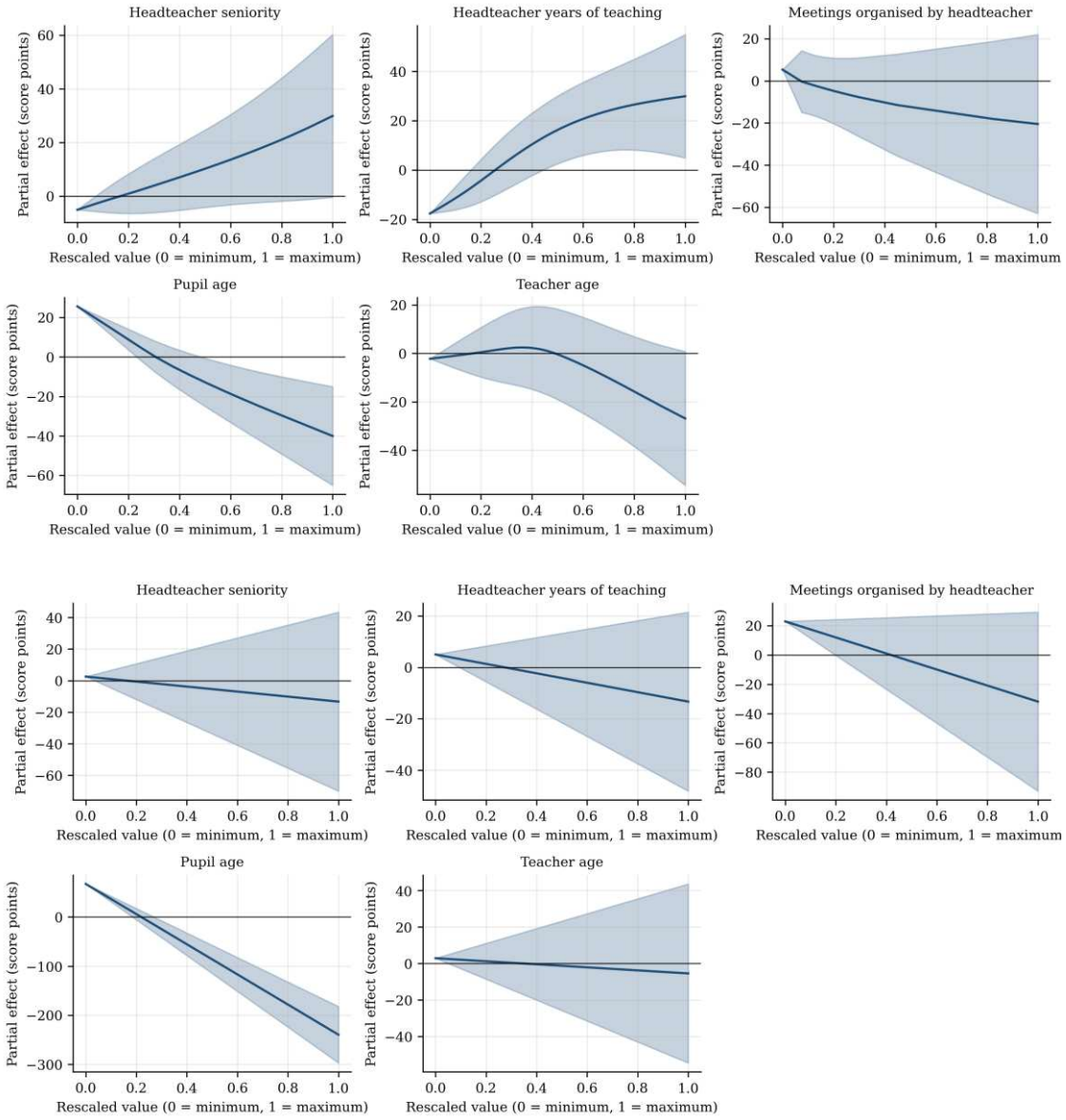


Figure 5. Partial effects of the five continuous variables in the additive model, with 95% bands: all schools, public schools, private schools. Source: author's calculations, PASEC 2019.

Variable	All: χ^2 (df), p	Public: χ^2 (df), p	Private: χ^2 (df), p
Headteacher seniority	10.3 (9), 0.329	14.3 (9), 0.113	0.6 (2), 0.755
Headteacher years of teaching	23.2 (9), 0.006	34.3 (9), <0.001	2.8 (3), 0.422
Meetings organised by headteacher	12.1 (7), 0.099	3.5 (7), 0.836	3.2 (2), 0.198
Pupil age	139.0 (9), <0.001	47.4 (9), <0.001	119.9 (2), <0.001
Teacher age	9.6 (9), 0.385	15.3 (9), 0.083	1.0 (3), 0.795

Table 8. Wald tests (school-clustered) on the spline terms of the additive model; degrees of freedom are the rank of the tested block, which is small in private schools because some variables take few distinct values. Source: author's calculations, PASEC 2019.

4.4 Discussion in light of the literature

Headteacher experience. The role of the headteacher's classroom experience agrees with the view that professional credibility supports instructional leadership (Leithwood et al., 2008) and with evidence that school-leader quality matters in African settings (Kadio, 2023, 2025). The absence of any robust role for seniority once teaching experience is accounted for, in all three samples, refines this reading: what is associated with learning is the experience of teaching, not the time in the post. Our contribution is to show that this association survives school-clustered inference and multiple-testing correction, which most earlier work does not apply, and that it is the only one to do so in the pooled sample.

Management quality. The association of management quality is in line with the cross-country result that better-managed schools perform better (Bloom et al., 2015) and with evidence that managerial capacity in public schools of low-income countries is weak and matters for outcomes (Anand et al., 2023). It is, however, fragile in our data: it is significant before correction in all schools, robust only at the 10% level in public schools, and absent in private schools.

Rewards. The association of rewards for the best pupils agrees with self-determination theory (Deci & Ryan, 2013) before correction, and the inverted-U decile profile in public schools is compatible with this reading without demonstrating it. After correction it is not distinguishable from noise, and reverse causation is plausible, since better schools may reward more.

Pedagogical practices and relations. Teacher inspection and pedagogical meetings are not robust, which is consistent with the finding that accountability tools produce effects only when they change incentives or practice (Bruns et al., 2011). For parental relations, the absence of any association ($\theta = -0.5$, $p = 0.864$ in all schools), with a negative OLS sign, is consistent with Hill and Tyson (2009): if contact is mostly reactive, an average near zero can hide effects of opposite sign; schools may also contact parents because pupils struggle.

Public and private schools. Robust associations appear in public schools only. This may reflect more homogeneous practices among private schools, or the smaller sample (179 training schools against 858) and wider intervals; our data cannot tell the two apart.

Relation to earlier work using the same data. Assienin et al. (2025) classified pupils with neural networks and concluded from Shapley values that administrative management, pupil relations and parental relations were systematic determinants. The present estimates differ: none of the three is significantly associated with the score in all schools, even before correction ($p = 0.164$, 0.381 and 0.864). The earlier ranking cannot test a hypothesis and rested on a split that does not separate schools. The two papers are complementary, but the earlier causal wording is not supported.

Methodological reading. The use of Holm-type corrections follows List et al. (2019), and machine learning is used here only for nuisance estimation within DML (Chernozhukov et al., 2018), with school-grouped folds as recommended for hierarchical data (Roberts et al., 2017). The comparison with tree-based learners shows that, for this problem, flexibility is not rewarded on unseen schools.

5. Conclusions

Only one governance association survives school-clustered inference and multiple-testing correction in the pooled sample: the headteacher's years of teaching, which is positively associated with learning at every decile of the score distribution, with a peak near the median. In public schools it is joined, at the 10% level, by the quality of school management, while no governance indicator survives in private schools. Management quality, rewards for the best pupils and the number of meetings organised by the headteacher are significant before correction in some samples and should be treated as suggestive only. Parental relations are not associated with learning in either direction. Linear models predict as well as additive models and tree-based ensembles on schools held out of estimation, and the observed covariates explain about 31% of the variance in all schools, 25% to 26% in public schools and less than 7% in private schools.

5.1 Implications for economic policy

Two findings can inform decisions, with the caveat that they are associations. First, the selection and training of headteachers could give weight to prior classroom experience, since it is the one indicator robustly associated with learning. Second, support for management quality, through training and supervision, is a candidate for public spending on education quality in public schools, where the association holds at the 10% level, but the evidence is weaker than for experience. Policies that treat the frequency of school–parent contact as an indicator to maximise are not supported; monitoring tools that distinguish preventive from reactive contact would be a better basis for evaluation. Evaluations with an identification strategy (randomised or staggered rollouts of headteacher training, say) are needed before any of these is scaled up. For private schools, recommendations should remain cautious.

5.2 Limitations

1. **No identification.** The data are cross-sectional and unobserved school characteristics may drive both governance and learning; all estimates are conditional associations.
2. **Plausible values and weights.** The mean of ten plausible values is treated as observed and sampling weights are not used, so standard errors are probably too small. Estimation by plausible value with Rubin's rules is the natural extension.
3. **Multiple testing.** The 10% Holm level was chosen with a family of sixteen tests in mind but after seeing results; at 5%, management quality in public schools no longer survives and only years of teaching remains in all and public schools. Both levels are reported.
4. **Training sample and power.** Estimation and inference use the 80% of schools of the training sample, so that the test sample stays untouched for prediction; intervals would be somewhat narrower on the full sample. In private schools, 179 training schools and 45 test schools give wide intervals and a test that cannot rank the models; null results are not evidence of no effect.
5. **Quantile inference.** Wald tests use an analytical covariance that ignores school clustering.
6. **Single split.** All predictive results rest on one split of schools; the cross-validated R^2 (with its standard deviation) is reported alongside to show the sensitivity to the choice of schools.
7. **Additive model in private schools.** The smoothing criterion keeps decreasing up to the end of the grid, so the model is effectively linear there, and some spline blocks have a low rank.

8. **Single wave.** Only PASEC 2019 is used.

6. References

African Development Bank Group. (2020). *African economic outlook 2020*. African Development Bank.

Anand, G., Atluri, A., Crawfurd, L., Pugatch, T., & Sheth, K. (2023). *Improving school management in low- and middle-income countries* (IZA Discussion Paper No. 16252). IZA.

Assienin, S. K., Kouakou, A. K., N'da, C. K., & Yobouet, L. L. E. (2025). School governance and learner performance in Sub-Saharan Africa: A neural networks approach. *International Journal of Economics and Finance, 17*(9), 90–105.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. National Bureau of Economic Research.

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

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., & Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa. *Journal of Economic Perspectives, 31*(4), 185–204.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Bruns, B., Filmer, D., & Patrinos, H. A. (2011). *Making schools work: New evidence on accountability reforms*. World Bank.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal, 21*(1), C1–C68.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science, 11*(2), 89–121.

Epstein, J. L., & Jansorn, N. R. (2004). School, family, and community partnerships link the plan. *Education Digest, 69*(6), 19–23.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877–896.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14*(3), 351–388.

Hanushek, E. A., & Woessmann, L. (2023). [Full reference to be completed from the thesis bibliography.]

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology, 45*(3), 740–763.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Kadio, K. (2023). *Education Economics, 31*(2), 166–188. https://doi.org/10.1080/09645292.2022.2052807 [full title to be completed from the DOI record]

Kadio, K. (2025). *International Journal of Educational Development, 115*. https://doi.org/10.1016/j.ijedudev.2025.103290 [full title to be completed from the DOI record]

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33–50.

Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association, 94*(448), 1296–1310.

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management, 28*(1), 27–42.

List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics, 22*(4), 773–793.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics, 29*(3), 305–325.

Moisset, P., et al. (2018). [Full reference to be completed from the thesis bibliography.]

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.

PASEC. (2020). *PASEC2019: Qualité des systèmes éducatifs en Afrique subsaharienne francophone*. CONFEMEN.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Roberts, D. R., Bahn, V., Ciuti, S., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929.

Torre, D., & Murphy, J. (2016). [Full reference to be completed from the thesis bibliography.]

World Bank Group. (2022). [Full reference to be completed from the thesis bibliography.]