Making School-Set Marks Comparable

Does Lenient Grading Buy a Better Secondary School?

Peter Courtney

Stellenbosch University and Vrije Universiteit Amsterdam

Presented at the VU Economics Lunch Seminar
Vrije Universiteit Amsterdam · 30 September 2026

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

The problem: why education systems need comparable marks

- Comparable testing is rare and retreating, so most primary pupils are assessed only by their own school
  - At most 25% of countries test every pupil at the end of primary; a third of 63 African and South Asian systems have no primary exam
  - In South Africa, the first comparable assessment is the school-leaving exam (NSC)
- **Why is this a problem?** Learning cannot be compared across schools.
  - Testing raises learning only where results compare across schools, because comparison holds schools to account (Bergbauer, Hanushek and Woessmann 2024)
  - Reforms are judged on proxies such as class size, not on what pupils learn
  - The rich–poor gap cannot be measured, nor where a district should direct support
  - Secondary admission and high-school value added rest on incomparable marks

The World Bank's 2018 review found "no effective way to aggregate … classroom-level formative assessment by teachers into … reliable system-level information".

This paper achieves exactly that, without a new test.

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

This paper

- South Africa has no common exam at the end of primary school: every Grade 7 mark is set by the school that awards it. This paper constructs a common scale, and asks how far each school diverges from it and what that divergence implies.

- Almost every pupil moves from primary to secondary school, from Grade 7 to Grade 8. Pupils from many primary schools converge on each secondary school, where they write a common set of assessments. This creates the first layer of comparability.

- The gap between the common scale and a school's marks measures how lenient or strict the school is. This paper asks whether marking leniently, awarding a higher mark than an average school would for the same learning, lets pupils attend stronger secondary schools.

Main findings

1. **Schools grade on widely different scales.** Schools at the 10th and 90th percentiles of strictness mark the same learning more than an SD of pupils' marks apart.
2. **Strictness is a trait of the school.** Estimated on two random halves of each school's pupils, the two estimates correlate at 0.88; across cohorts four years apart, at 0.60.
3. **Marks on the common scale predict the school-leaving exam better.** Five years on, they beat raw marks against the NSC in each of the six provinces, in both Mathematics and Mathematical Literacy.
4. **Raw marks misrank schools and understate the rich–poor gap.** 65.8% of schools change learning quintile. The poorest schools grade most leniently, so raw marks credit the richest schools with less than half their advantage.
5. **A lenient mark is discounted at the move to secondary school.** Holding her own mark fixed, a pupil from a primary one SD stricter joins Grade 8 peers scoring 0.14 SD higher. On its own, her primary's raw average predicts almost none of it.

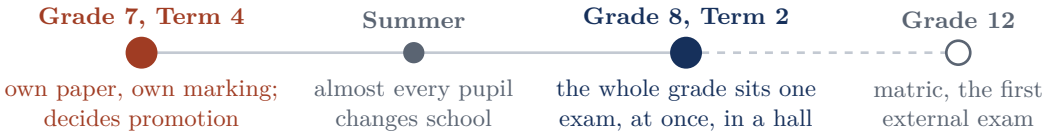
Related literature

- Comparable marks need a test set outside the school: a central exam at the end of secondary, or a researcher's own test of a sample of pupils. **Here the outside test already exists: each secondary school's own Grade 8 exam.**
  Schwerdt and Woessmann 2017; Woessmann 2003; Singh 2015; Araujo et al. 2016; Bau and Das 2020; Andrabi et al. 2011
- The estimator comes from matched employer–employee methods and its uses in education. **Its usual concerns, connectedness and limited-mobility bias, bind less here: most of the cohort changes school.** → A1
  Abowd et al. 1999; Mansfield 2015; Andrews et al. 2008; Kline et al. 2020
- Grading distortions are studied at thresholds, as marks piled just above a cut-off. **A school's strictness shifts every mark it awards, not only those near a cut-off.**
  Diamond and Persson 2016; Dee et al. 2019
- Households respond to published test scores, and choose schools for their peers rather than their effectiveness. **Below the school-leaving exam nothing is published, yet households sort as if on the calibrated scale.**
  Hastings and Weinstein 2008; Abdulkadiroğlu et al. 2020

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

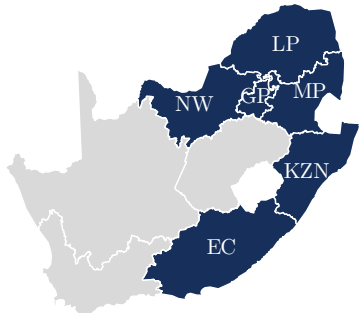
Institutional setting



- **Promotion cut-offs at 30, 40 and 50%.** Parents, the department and the cost of holding pupils back all push marks up.
- **Marks go to the department.** Teachers capture them; the principal signs off.
- **Parents see only their own child's marks** and a high school's matric results.
- **Mathematics only.** English failed the reliability screen.

The data

- **SA-SAMS**, South Africa's school administration system, holds each pupil's records, linked across years by a learner identifier. Term marks are graded from 0 to 100, with assessments set and graded at the school level.
- SA-SAMS records school-based assessments (SBAs), which teachers capture at the school; it is the system that produces every pupil's report card.
- The NSC is the National Senior Certificate, the high-school completion assessment, set and graded nationally.
- The primary analysis cohorts began Grade 7 in 2018 and 2022, followed one year on into Grade 8 in 2019 and 2023 respectively.
- The analysis from Grade 7 to the NSC follows the 2020 Grade 7 cohort until it writes the NSC in 2025.



The extract covers six provinces.

Data and attrition

	2018 cohort		2022 cohort	
Step	**N**	**%**	**N**	**%**
Total Grade 7 enrolment in SA-SAMS	796,928		866,679	
With valid Grade 7 Term 4 maths mark	708,659	88.9	804,654	92.8
Successfully linked to a Grade 8 record	544,233	68.3	674,557	77.8
Not linked: repeated Grade 7 the next year	—		27,339	3.2
Final analysis sample	532,295	66.8	665,019	76.7

Sample: national 2018 and 2022 cohorts; each percentage is of the cohort's total Grade 7 enrolment in SA-SAMS, the schools' administration system; the repeaters count exists for 2022 only (—); the Gauteng estimation sample is roughly 5 times smaller.

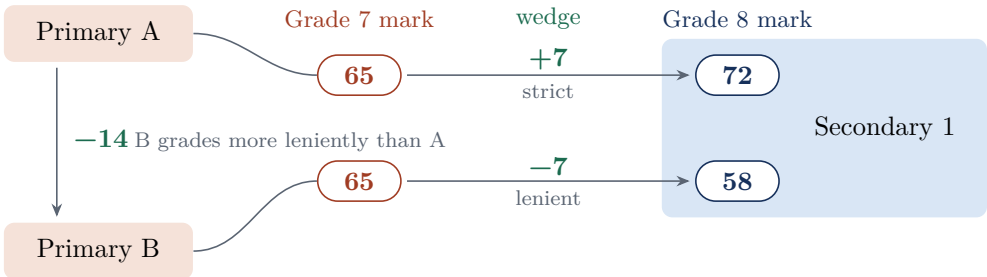
- Linkage is close to orthogonal to the school measure: the correlation with the calibrated level is 0.021, and an inverse-linkage-probability estimator shifts the mean by -0.005 marks.

Roadmap

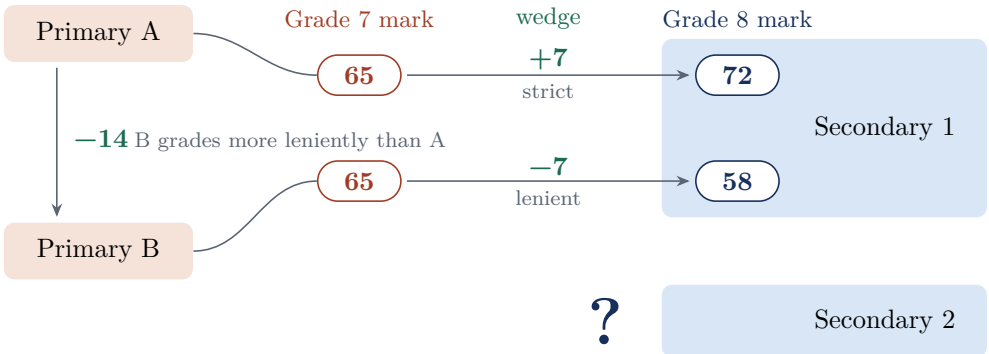
1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

Converging to a secondary school makes two primary schools comparable

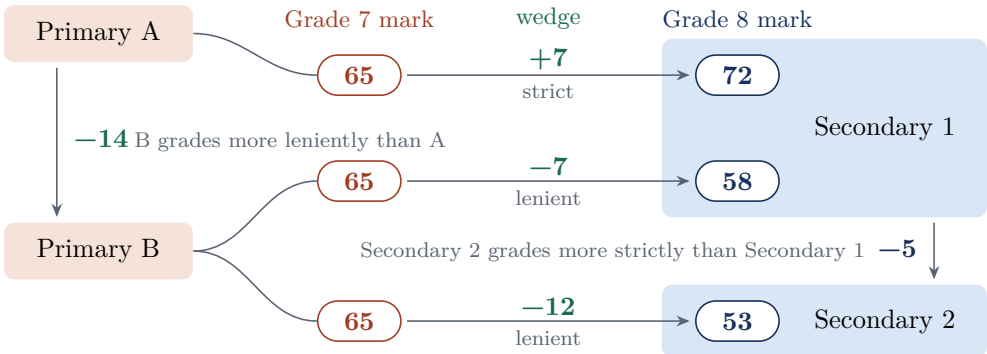
Same mark in, different learning out. The gap is each primary's grading wedge.



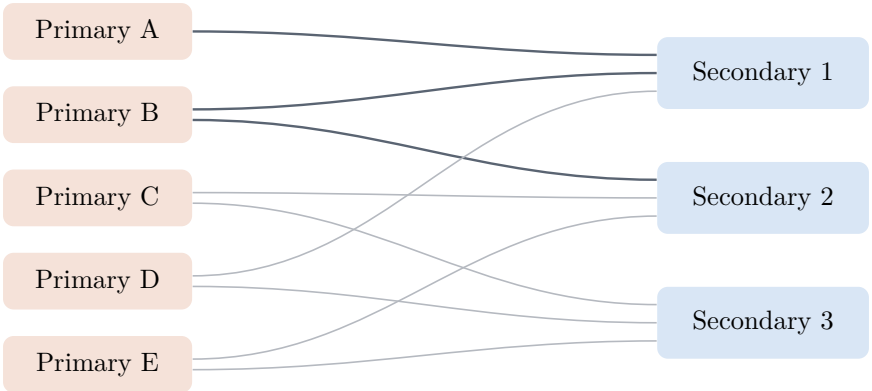
How do we compare additional secondary schools?



Diverging from a primary school makes two secondary schools comparable



The overlapping network chains all schools onto the common scale



The estimating equation

→ A2

Latent model

$$a^7_{ij} = g^7_{ij} + s_j + \varepsilon^7_{ij} \tag{1}$$

- a^7_{ij}: learning
- g^7_{ij}: recorded mark
- s_j: primary j's grading offset, > 0 if strict

$$a^8_{ijk} = g^8_{ijk} + d_k + \varepsilon^8_{ijk} \tag{2}$$

- d_k: secondary k's grading offset

$$a^8_{ijk} = c + \lambda\, a^7_{ij} + \nu_{ijk} \tag{3}$$

- c: six-month growth
- λ: persistence

⇓ substitute

Estimating equation

$$g^8_{ijk} = \underbrace{\beta_0}_{c} + \underbrace{\beta_1\, g^7_{ij}}_{\lambda} + \underbrace{\alpha_j}_{\lambda s_j} + \underbrace{\gamma_k}_{-d_k} + \varepsilon_{ijk} \tag{4}$$

α_j: the wedge

Identifying assumptions

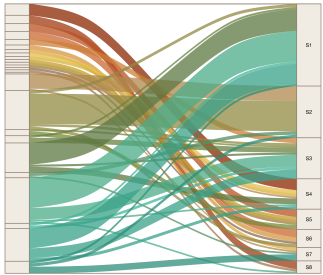
The assumption: pupils with the same learning at the end of primary get the same Grade 8 mark at the same secondary, whatever their primary. Plausible, because the marks are months apart.

	Breaks it if	**In the data**
Leniency is a shift	A primary moves its strong and weak leavers' marks by different amounts.	Holds in level, not in slope: the gap widens with the mark. Allowing for it, ranks correlate 0.996. → A3
A single scale per secondary	A secondary marks some primaries' entrants differently, or a primary teaches to its tests.	Small where testable: net of noise, a pair of schools departs from their sum by 0.08 SD. Teaching to the test is maintained.
No sorting on growth	Leavers on the same mark pick a secondary by how much they will grow: summer loss, early Grade 8 learning.	Maintained. Simulated sorting 2.3 times what the data show would bias the wedge's variance by 5%. → A4

It also needs every school linked into one network: the next frame.

The transition network

A representative region of primary (LHS) to secondary (RHS) transitions



6
secondary schools receive the leavers of the median primary

10
primary schools feed the median Grade 8 class

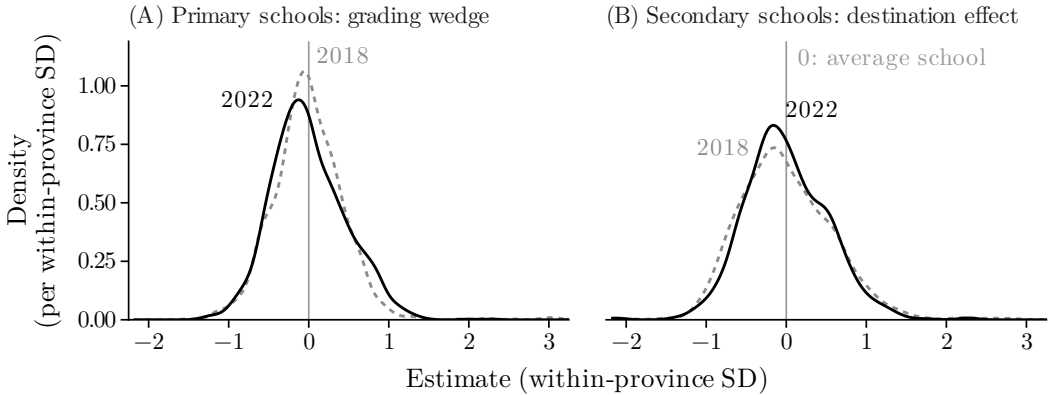
98.8%
of schools sit in the connected set in the weakest province-cohort cell, and more in every other

- **Limited mobility does not bind.** With no pupil fixed effect, all of a primary's linked leavers identify its wedge: 40 at the median primary.
- **Nor does a thin network.** Gauteng's algebraic connectivity is 14 (2018) and 12 (2022) times the Jochmans–Weidner weak-connectivity benchmark.

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

The wedge and destination effect distributions



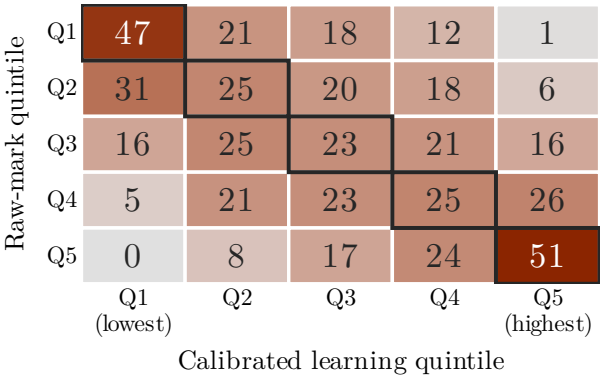
α is relative: zero is the average primary school in the province and cohort.

Sample: Gauteng mathematics, 2018 and 2022; 1,424 and 1,479 primary, 901 and 839 secondary schools.

The effect size is substantial: two-thirds of schools change quintile

65.8% of schools change quintile once the wedge is removed

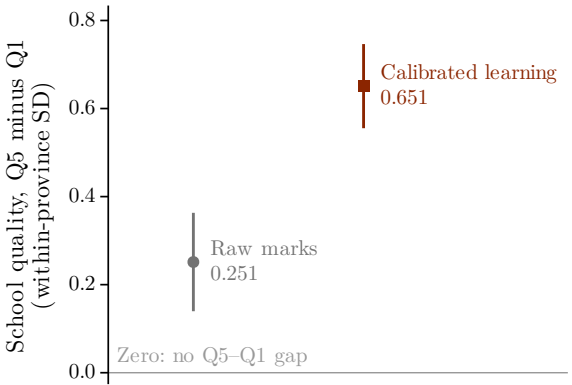
Learning quintile before and after calibration



- **Even the extremes turn over:** only 47% of bottom-quintile and 51% of top-quintile schools stay put.
- The middle quintile keeps 23%; the rest spread almost evenly across all five.
- Most movers shift one quintile, but 28% of all schools move two or more.

The recalibration is important: the SES gradient more than doubles

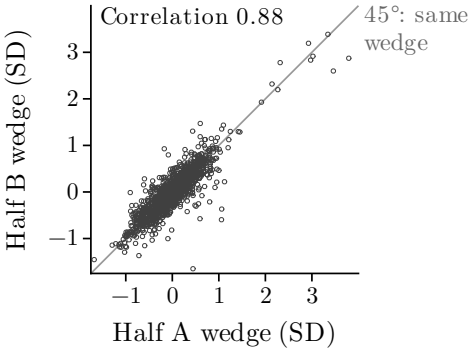
One school-level regression, run twice: raw mean Grade 7 mark, then calibrated learning level. Same controls both times.



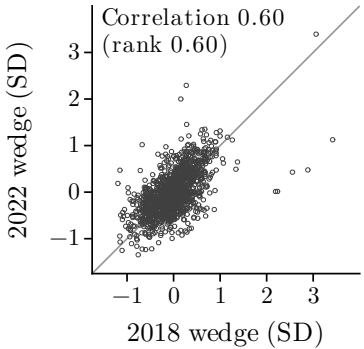
- Quintile 5 leads Quintile 1 by 0.251 SD on raw marks and 0.651 SD on calibrated learning; the bootstrapped intervals do not overlap. → A12
- No-fee schools grade about 0.47 within-province SD more leniently than fee-charging schools; no-fee status and the school's own poverty quintile are the only resource measures surviving Romano–Wolf. → A11
- **Raw marks credit the richest schools with less than half their advantage.**

The wedge is a property of a school, not a learner cohort

Two random halves of each school, 2018

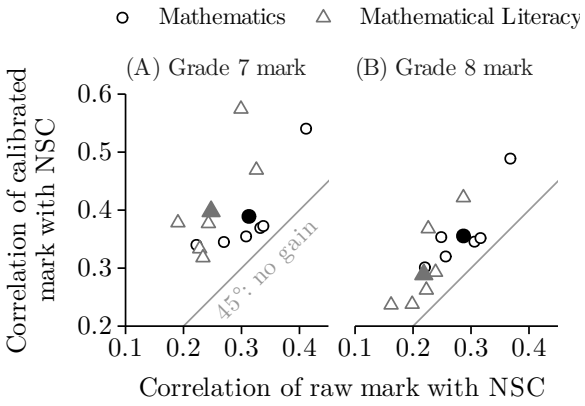


Two cohorts, four years apart



New pupils, new tests, no link between the years: what persists belongs to the school.

Validation: five years later, the NSC



- Cohort: Grade 7 in 2020, six provinces. Outcome: their NSC mark in 2025, **five years later**
- Calibrated learning predicts it better than raw marks in 14 of 14 fits (two subjects, six provinces and pooled), at Grade 7 and again at Grade 8
- Grade 7: adding the primary school's grading wedge raises the correlation from 0.31 to 0.39
- Grade 8: removing the high school's offset raises it from 0.29 to 0.36
- The weights come from the Grade 8 calibration and were never fitted to the NSC

Each point is one subject × province correlation with the NSC mark; filled points are the six provinces pooled.

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

Does lenient grading buy a better secondary school?

- Primary A marks strictly, Primary B leniently. A pupil leaves each with 65.
- If secondary schools read 65 at face value, both join equally strong classes. If the lenient 65 is discounted, B's pupil joins weaker ones.
- **The test.** Hold her own mark fixed: does her primary's wedge predict how her new classmates do in Grade 8? Classmates from her own primary are left out.
- **Scope:** this is not a causal estimate of what generates the observed sorting.
  → A10

At the same mark, pupils from lenient primaries end up among weaker classmates

- Her new classmates' Grade 8 achievement, her own mark held fixed, regressed on her primary's **raw Grade 7 mean** (row 1) or its **grading wedge** (row 2). LOFO drops classmates from her own primary.

	Her new classmates' Grade 8 achievement	
Origin signal	**Full**	**LOFO**
Raw Grade 7 mean	0.023 (1.6)	**0.020 (1.4)**
Grading wedge $\hat{\alpha}_j$	0.165*** (9.2)	**0.136*** (7.4)**

Raw average: 0.020

On their own, her primary's raw marks do not predict how strong her classmates are.

Grading wedge: 0.136*

A primary one SD more lenient: classmates 0.136 SD weaker, none from her own primary. → A9

Value added: the calibrated mark as an intake

Value added: how far a school's students finish above what their intake predicts

$$y_{ik}^{12} = \theta_k + \delta\left(\hat{\beta}_1\, g_{ij}^{7} + \hat{\alpha}_j\right) + \varepsilon_{ik}$$

- $\hat{\alpha}_j$: primary j's wedge: puts the intake on the common scale
- g_{ij}^{7}: Grade 7 mark, set by primary j on its own scale
- θ_k: high school k's value added
- y_{ik}^{12}: school-leaving (NSC) mark, Grade 12

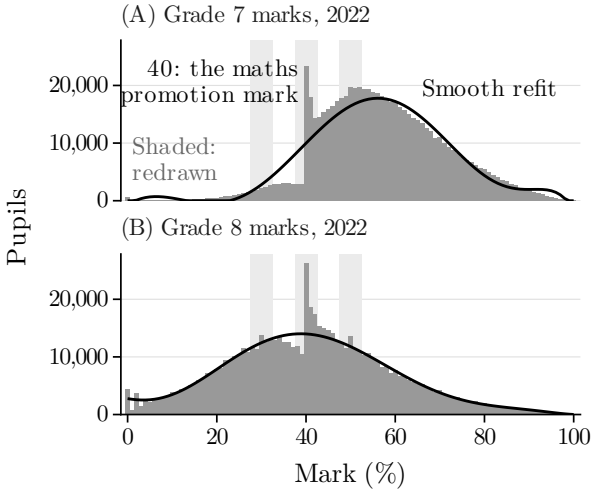
Without the wedge, a high school gets credit for its feeders' strict grading.

40% of high schools change quintile of value added when the intake is calibrated.

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

The impact of threshold pushing



- **Marks pile up at 40**, the maths promotion mark: eight times as many Grade 7 pupils on 40 as on 39
- **Aggressive by design:** marks within two points of 30, 40 or 50 are redrawn from the curve, 25% of Grade 7 and 31% of Grade 8. Not the true distribution
- **The wedge does not move:** refitted and raw wedges correlate at 0.997 or more in every province and cohort
- **The pushers are not the lenient schools:** the strictest fifth put 2.1–3.4 times as many marks on the thresholds as the most lenient fifth (Gauteng)

Roadmap

1. Motivation
2. This paper
3. Setting and data
4. Methodology
5. Results
6. Applications
7. Robustness
8. Close

Conclusion

By exploiting the transition from primary to secondary school, this method can recover comparable learning, without a new test, in the 75% of countries whose primary school results are not comparable. From this follow high-school value added, evaluations of what primary schools do, and the rich–poor gap measured on the common scale.

Robustness at a glance

Threat	Check	What moved
Sorting on unobservables	Simulated sorting and endogenous mobility	Levels stretch or shrink; ranks hold at 0.897 or better → A4
Differential decay by achievement or poverty	Grade 7 slope by achievement and by poverty quintile; Grade 8 anchor moved from Term 1 to Term 4	No-fee gap keeps 99–101% of its size; Term 1 and Term 4 wedges correlate 0.90 → A5
Leniency changes with the mark	Separate slopes for strict and lenient primaries, within secondary	Holds in level, not in slope; ranks 0.996 → A3
Thin network, limited-mobility bias	Connectivity against the Jochmans–Weidner benchmark	Nothing: connectivity is 14 (2018) and 12 (2022) times the benchmark → A1
Functional form or estimator misspecified	Quadratic, piecewise-linear, empirical-Bayes, cross-classified	Wedges correlate 0.987–0.998 with the baseline → A6
Second stage runs on a noisy estimated wedge	Paired cluster bootstrap, wedge re-estimated per replicate	Interval widths within 7% of the naive → A12

Appendix index

Identification and sorting

- **A1** Network density and limited-mobility bias
- **A2** The latent model in full
- **A3** Leniency shifts the level, and the slope differs too
- **A4** Sorting on growth: the simulated bound and its limit
- **A5** Differential decay: the no-fee gap holds its size
- **A6** Four other specifications give the same wedges
- **A13** Where does pupil ability enter?

Sample and measurement

- **A7** The $\lambda\rho$ decomposition

The sorting test's own robustness

- **A8** Joint specification and six-province panel
- **A9** The sorting result survives every de-leaking
- **A10** The geography bound

Downstream and comparative

- **A11** What the wedge is made of
- **A12** All eight coefficients, raw and calibrated

A1 · Network density and limited-mobility bias

- Harmonic-mean weighted degree 202.9; the giant component covers 99.6% of schools pooled across the whole network.
- Algebraic connectivity $\lambda_2 = 0.0537$ (2018) and 0.0474 (2022), about 14 and 12 times the Jochmans–Weidner weak-connectivity benchmark of 0.0039 at which fixed-effect standard errors become severely understated.
- School networks are denser than matched employer-employee panels for a structural reason: primaries outnumber secondaries, and every pupil moves in the same year.
- **Answers "is this AKM?".** The leave-out correction of Kline, Saggio and Sølvsten puts estimation noise at about 5% of the wedge's variance: small, not nil.
- **No minimum link size.** Dropping links under five pupils adds noise (3.6% to 16.6% of the variance, 2018); net of it, ranks hold (0.95, 2018; 0.99, 2022).

← Appendix index

A2 · The latent model in full

$$a^7_{ij} = g^7_{ij} + s_j + \varepsilon^7_{ij}$$
$$a^8_{ijk} = g^8_{ijk} + d_k + \varepsilon^8_{ijk}$$
$$a^8_{ijk} = c + \lambda a^7_{ij} + \nu_{ijk}$$

Substituting and relabelling the structural composites gives

$$g^8_{ijk} = \beta_0 + \beta_1 g^7_{ij} + \alpha_j + \gamma_k + \varepsilon_{ijk}$$

with $\beta_0 \equiv c$, $\alpha_j \equiv \lambda s_j$, $\gamma_k \equiv -d_k$, $\beta_1 \equiv \lambda$. That β_1 is structural. The *estimated* one is $\lambda\rho$, and only if the first line is read as measurement error **in the mark**:

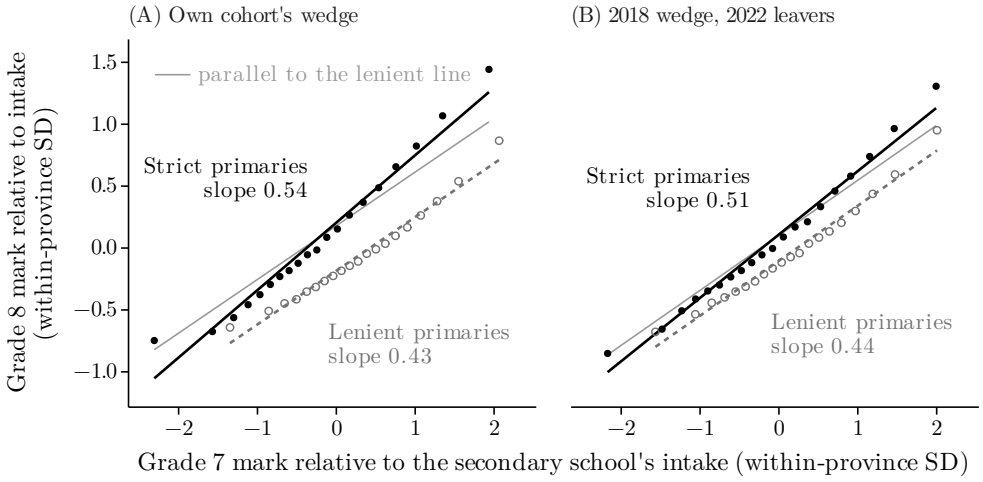
$$g^7_{ij} = a^7_{ij} - s_j - \varepsilon^7_{ij}, \qquad \varepsilon^7_{ij} \perp a^7_{ij}$$

so ε^7 sits inside the regressor and $\rho = \mathrm{Var}(\tilde{a}^7)/\mathrm{Var}(\tilde{g}^7)$ is the mark's reliability.

- s_j and d_k are the same object, an additive school-set offset between true learning and the recorded mark. A positive offset means the mark sits below true learning, so $s_j > 0$ is a strict origin.
- They enter asymmetrically for one reason: position in time. The origin offset is scaled by persistence because it distorts the Grade 7 mark on the right-hand side; the destination offset distorts the outcome directly.
- Measurement error in g^7 puts the slope's probability limit at $\lambda\rho$, so λ and ρ are not separately identified from the cross-section, and each is bounded below by the estimated slope.
- It reaches $\hat{\alpha}_j$ too: with $\hat{\alpha}_j \approx \overline{g^8}_j - \beta_1 \overline{g^7}_j$, fitting β_1 at $\lambda\rho$ leaves $\hat{\alpha}_j$ carrying a term in the school's own mean Grade 7 mark. The contamination is **correlated with the raw mark**, not shrinkage toward zero.

← The estimating equation

A3 · Leniency shifts the level, and the slope differs too



← Appendix index

A4 · Sorting on growth: the simulated bound and its limit

- **What breaks the wedge.** Leavers of one primary, on one Grade 7 mark, reaching different secondaries because of what they will learn next. Sorting on the mark, a primary's wedge or a secondary's quality is absorbed by $\hat{\beta}_1 g_{ij}^7$ and the fixed effects.
- **The simulation.** Within each primary, pupils of higher unobserved ability reach better secondaries, at strength ψ; each draw is fitted with the paper's two-way regression.
- **The footprint in the data.** Within a primary, a leaver's Grade 7 mark and her secondary's effect correlate at 0.021: in the simulation, $\psi \approx 0.05$, which moves the wedge's variance by 2.3 percentage points.
- **The threshold.** A 5-point move needs a footprint of 0.048, 2.3 times the data's. At $\psi = 2$, far past it, ranks still correlate 0.897 with the truth while the variance shrinks to 0.73 of its true size.
- **What it cannot see.** Sorting on growth unrelated to the mark leaves no footprint. The assumption is maintained; the simulation sizes only sorting that travels with the mark.

← Appendix index

A5 · Differential decay: the no-fee gap holds its size

- **The threat.** If learning faded faster over the holiday for poorer or weaker pupils, their primaries would book as lenient graders.
- **A slope for each group** (Gauteng, paper appendix H.8). Re-estimating the wedge with a Grade 7 slope for each poverty quintile, or for each quintile of the school's mean Grade 7 mark, leaves the no-fee gap at 0.99–1.01 of its single-slope size.
- **A later anchor.** Ongoing fade would widen the gap as the Grade 8 anchor moves later in the year. Anchored on Terms 2, 3 and 4, it is 0.50, 0.48 and 0.51 of the wedge's SD.
- **Term 1 against Term 4** (national extract). Wedges anchored on the two terms correlate 0.90, precision-weighted, over 6,676 primaries.
- **What it cannot remove.** Within a secondary, no-fee pupils' Grade 8 marks rise less with their Grade 7 mark: the interaction is −0.149, about 0.27 of the fee-school slope. That is A3's slope component; the design cannot say whether it is grading or learning.

← Robustness at a glance

A6 · Four other specifications give the same wedges

Specification	What changes	Correlation
Quadratic	Adds the squared Grade 7 mark	0.994
Piecewise linear	A kink at the median Grade 7 mark	0.998
Empirical Bayes	Shrinks each wedge to the cohort mean	0.998
Cross-classified	Crossed random intercepts for both schools	0.987

The correlation is with the baseline wedge. Each specification changes the common Grade 7 to Grade 8 relation or the estimator and keeps one offset per school, so none can test the shift; A3 does.

← Robustness at a glance

A7 · The $\lambda\rho$ decomposition

- The slope bounds both persistence and reliability from below: each lies in $(0, 1]$ and their product is the estimated slope, so persistence is at least 0.59 and the Grade 7 mark is at least 59% signal in both cohorts.
- A test-retest and instrumental-variables decomposition brackets the 2018 cohort at reliability $\rho \approx 0.70$ to 0.71 and persistence $\lambda \approx 0.75$ to 0.85, so attenuation likely dominates.
- Nothing in the paper's strategy depends on the split. What matters is the direction: with $\rho < 1$ the wedge is attenuated toward zero, so every reported magnitude is a floor.

← Appendix index

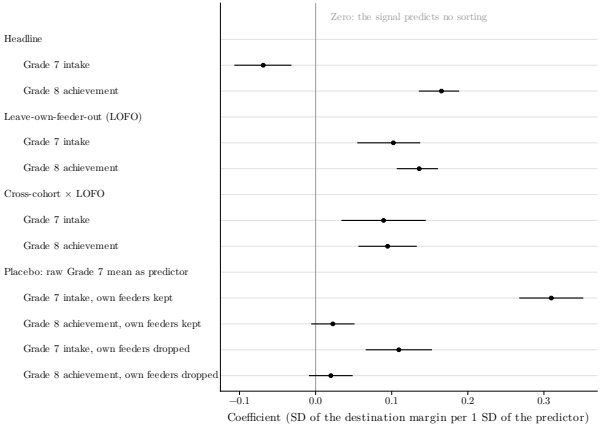
A8 · Joint specification and the six-province panel

Province	n	Full	LOFO	t
EC	162,355	0.068	0.013	0.8
GT	220,973	0.166	0.135	7.2
LP	182,360	0.107	0.043	2.5
MP	106,955	0.144	0.091	4.3
NW	91,119	0.117	0.047	2.4
ZN	254,501	0.068	0.013	0.8
POOLED6	1,018,263	0.117	0.064	8.4

- The Gauteng focus is a precision choice, made for the network-density reason given in the data section, not a result-shopping one. The magnitude roughly halves when the six provinces pool and stays clearly positive.
- The panel is six provinces rather than nine because the extract covers six; a move into the remaining three is recorded as attrition.
- Entering all three origin signals jointly is a bad-control exercise: the raw Grade 8 mean contains the wedge by construction, so the joint fit splits one object across two regressors and the wedge's coefficient flips. Fitted without it, the raw Grade 7 mean and the wedge both predict stronger classmates, in the ratio sorting on the calibrated scale implies.

← Appendix index

A9 · The sorting result survives every de-leaking



- Substituting the wedge estimated on the *other* cohort, which severs estimation overlap, holds the achievement margin at +0.108; cross-cohort with the own-feeder cut at +0.095; netting out destination leniency through γ_k as well, all three severed at once, at +0.136.
- **The raw mark predicts intake strongly and achievement not at all.** The placebo dots are not all at zero: the raw Grade 7 mark on Grade 7 peers is the largest point on the chart. Only the Grade 8-outcome placebos sit at zero.

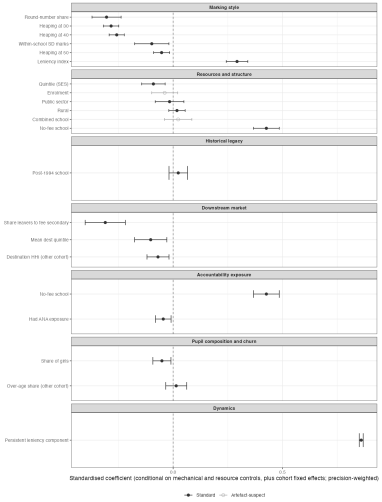
← At the same mark

A10 · The geography bound

- Among cross-ward switchers the achievement margin holds at 0.143 ($t = 7.7$); among cross-municipality switchers at 0.082 ($t = 3.0$). 64.1% of movers cross a ward and 5.7% cross a municipality.
- The distance interaction is -0.203 ($p < 0.01$): the margin attenuates with distance moved, but stays positive across the range.
- What this does **not** settle: residential relocation stays observationally equivalent to school choice. The paper claims only that enrolment tracks the wedge.

← Does lenient grading buy a better secondary school?

A11 · What the wedge is made of

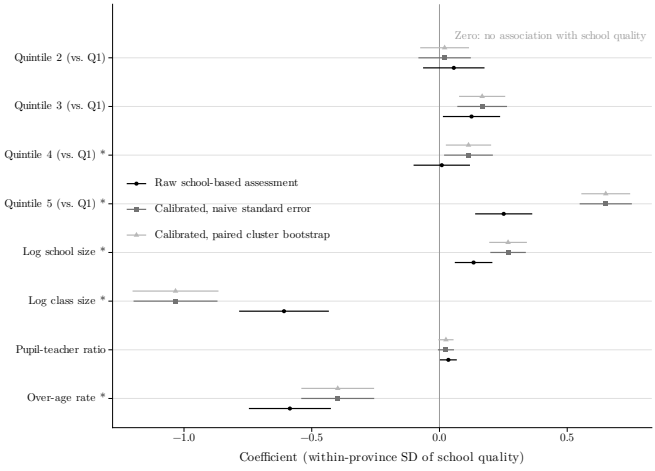


Family	Share of joint R^2
F6 Downstream market	38.3%
F3 Resources and structure	28.8%
F7 Accountability exposure	28.4%
F8 Pupil composition and churn	2.6%
F4 Historical legacy	1.8%

- Three families carry 96% of the joint R^2 of 37.1%, and downstream market leads without commanding.
- No-fee schools grade about 0.47 within-province SD more leniently than fee-charging schools; no-fee status and the school's own poverty quintile are the only resource measures surviving Romano–Wolf.
- Each 1 SD of destination-market concentration associates with 0.033 SD stricter grading, own SES held fixed.
- **Every association here is descriptive.** Four readings fit the market result equally well: precision mechanics, reverse causation, common cause, and repeated-game accountability. The design separates none of them.

← SES gradient

A12 · All eight coefficients, raw and calibrated



An asterisk on a row marks a term whose raw and calibrated coefficients differ at 5%.

← Appendix index

A13 · Where does pupil ability enter?

- **Through the Grade 7 mark.** It stands in for the pupil's own effect, and every linked leaver identifies the primary's wedge whichever way ability is modelled.
- **A pupil effect from all eight term marks** orders primaries as the wedge built on the average of the four Grade 7 terms does: Spearman 0.97 in 2018 and 0.96 in 2022.
- **It departs from the Term 4 wedge in 2018 only.** Every term's mark has a part of its own at each primary, about 0.43 SD for the 2018 Term 4 mark; the 2022 Term 4 mark is in effect a year mark, so its own part is 0.16 SD.

← Appendix index