

Common Currency

Making school-set grades comparable, and sorting on the exchange rate of learning

Peter Courtney

Stellenbosch University and Vrije Universiteit Amsterdam

Stellenbosch University RESEP BBL · 01 September 2026

Roadmap

1. Motivation
2. This paper
3. Related literature
4. Setting and data
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

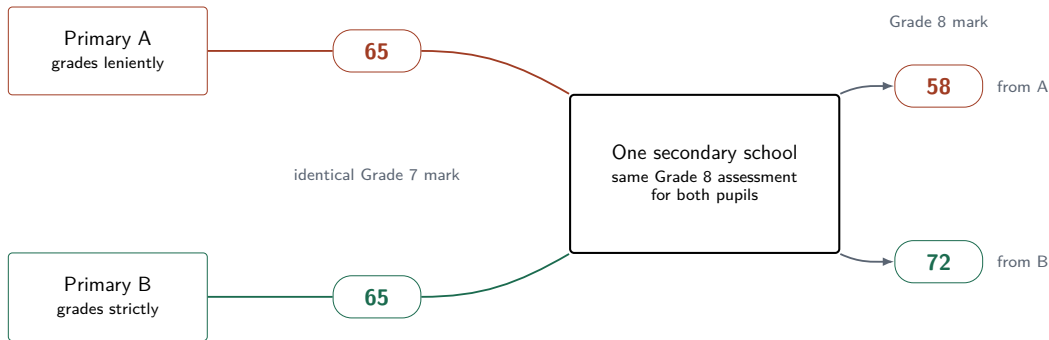
Below matric, every mark is set by the school that awards it

- The World Bank's flagship review says to start with "more and better measurement of learning". The binding word is *better*. Testing raises learning only where results are comparable between schools: by a quarter of a standard deviation, and by twice that in the weakest systems.
- Comparable measurement is rare and retreating. At most a quarter of the world tests every pupil on a common instrument at the end of primary, and 21 of 63 African and South Asian systems hold none. The exam was a gate: it rationed scarce secondary places. Where secondary supply has expanded there is nothing left to ration, so the gate goes and the only comparable measure goes with it. The systems dropping it are not the most developed.
- So in at least three quarters of the world a primary mark is quoted in the currency of the school that awarded it. South Africa's first external mark is matric (the school-leaving examination), five years later; every mark before it, including the Grade 7 mark used here, is school-set and school-marked.

The same review concluded there is "no effective way to aggregate ... classroom-level formative assessment by teachers into ... reliable system-level information". This paper does it.

The idea in one picture

Two pupils leave different primaries with the same Grade 7 mark and arrive at the same secondary school.



Same mark in, different learning out: Primary B graded more strictly.

Roadmap

1. Motivation
2. This paper
3. Related literature
4. Setting and data
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

What the wedge is, and what the common scale is

- **The Grade 8 anchor** is the assessment a secondary school sets for its whole grade. Everyone who walks through that door sits it, so it is the device that makes two primaries comparable. It is not the scale.
- **The grading wedge** $\hat{\alpha}_j$ is what the comparison recovers: how far a primary's marks sit below or above the learning behind them.
- **The common scale is the mark with the wedge added back**, $\beta_1 g_{ij}^7 + \hat{\alpha}_j$; averaged over a school's pupils, its calibrated learning level.
- **Strict** ($\hat{\alpha}_j > 0$): marks sit below the learning behind them, so leavers outperform their Grade 7 marks. **Lenient** ($\hat{\alpha}_j < 0$): the reverse. Neither word judges teaching.

Every wedge and learning-level magnitude here is in within-province standard deviations of the pupil mark distribution: one unit is the gap between the median pupil and the 84th percentile. The market section's coefficients are the exception, in within-cohort SD of the outcome per within-cohort SD of the signal.

What this paper does

- **Derives a common currency of learning** from administrative records on 1.2 million Grade 7 to Grade 8 transitions, in two cohorts, 2018 and 2022.
- **Prices every primary school against it**, recovering a grading wedge $\hat{\alpha}_j$ for each school in the connected set, at full administrative coverage and no marginal cost.
- **Shows the mispricing is large, persistent, structured, and that assortative sorting into secondary school aligns with the calibrated scale rather than with raw marks.** Leavers from a stricter primary end up in stronger Grade 8 cohorts; the raw school average families are shown predicts almost none of that.

Scope, stated up front. This is not a value-added estimate. $\hat{\alpha}_j$ says how far a school's marks sit from its pupils' measured learning; it does not say what the school contributed to that learning. Grading practice, teaching, peers and who enrolls can all move it, and the paper makes no attempt to separate them.

Preview of the main findings

1. **The mispricing distorts applied work.** 65.8% of schools change learning quintile once the wedge is removed; the SES gradient in school quality more than doubles.
2. **The exchange rate is signal, not noise.** Kane–Staiger reliability 0.97; split each school's pupils at random and the two halves rank schools the same way, at 0.88.
3. **Families already sort on it.** Hold a pupil's own Grade 7 mark fixed. A pupil leaving a primary that grades 1 SD stricter joins a Grade 8 cohort scoring 0.136 SD higher, on marks written by pupils who never attended her primary. Her primary's raw Grade 7 average, the only number families are ever shown, predicts 0.020 of that, which is not distinguishable from zero.
4. **The gap is a durable, structured school trait.** It persists at $r = 0.60$ across four years; the poorest schools grade most leniently.

Roadmap

1. Motivation
2. This paper
- 3. Related literature**
4. Setting and data
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

Grade inflation and manipulation at thresholds. (Diamond and Persson 2016; Dee et al. 2019; Jacob and Levitt 2003; Bertoni et al. 2013) Graders respond to incentives attached to particular grades, visible as bunching at a cut-off. The construct here is broader: a school-level stringency embedded in every grade a school awards, across the whole distribution.

School choice and published quality signals. (Hastings and Weinstein 2008; Abdulkadiroğlu et al. 2020; MacLeod and Urquiola 2015, 2019; Andrabi et al. 2017; Figlio and Lucas 2004; Black 1999) Those designs lean on a visible signal: a report card, a league table. Below matric nothing is published here, so a family's best signal is their child's Grade 7 mark, denominated in the school's own currency.

Two-way fixed effects with movers. The estimator comes from the matched employer-employee literature (Abowd et al. 1999; Card et al. 2013; Andrews et al. 2008; Kline et al. 2020; Jochmans and Weidner 2019) and its education applications (Kane and Staiger 2008; Chetty et al. 2014; Mansfield 2015). The tradition's headline concerns, network connectedness and limited-mobility bias, carry over. → [B11](#)

Measuring school quality where standardisation is scarce. (Singh 2015; Bau and Das 2020; Araujo et al. 2016; Bold et al. 2017; Andrabi et al. 2011) Administrative grades are often the only learning measure at scale.

What is new here

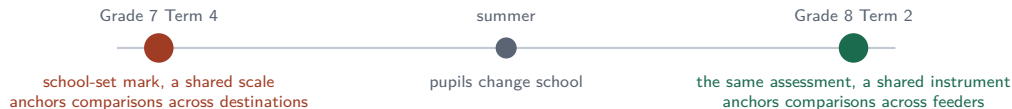
- **A price, not a production function.** Value-added asks what a school added, and uses controls to separate that from what its pupils brought with them. This asks a smaller question: what is a mark from this school worth? Nothing is controlled away. A strict school's mark converts upward whether it marks hard, teaches well, or enrolls well. → [B10](#)
- **Secondary schools do the linking that the assessment system never does.** Two pupils from different primaries who land in the same Grade 8 classroom write the same paper, so their primaries can be compared through them. Chain those comparisons across the network and every primary lands on one scale. No common test across destinations is required.
- **Full administrative coverage, at zero marginal cost.** Every school in the connected set is priced from records the system already holds, not from a sampled survey.

The closest design is Liu et al. (2025), which also anchors school effects on the institutions its pupils enter next. Their pupils sit a state test as they leave middle school, so the comparing is already done and the anchor buys them an attribution: value added. Their design begins where this one ends, with every pupil already on one scale. Here no two schools at either stage share an assessment, so the anchor has to do the comparing, and only the comparing. → [B10](#)

Roadmap

1. Motivation
2. This paper
3. Related literature
- 4. Setting and data**
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

Institutional setting: the Grade 7 to Grade 8 transition



- Grade 7 ends primary school, and nearly every pupil changes institution.
- Grade 8 Term 2: the whole grade sits a proctored common examination, and the recorded mark blends that examination with common continuous assessment.
- Almost every school sets **the same set of assessments** for its whole Grade 8 grade, examination and continuous-assessment tasks alike, so leavers from different feeders are marked against one another on identical work. No two destinations need share anything.
- Anchoring runs both ways: the destination supplies a shared **instrument**, and a feeder, grading all its leavers alike, supplies the shared **scale**.
- Mathematics only. English failed the reliability screen.

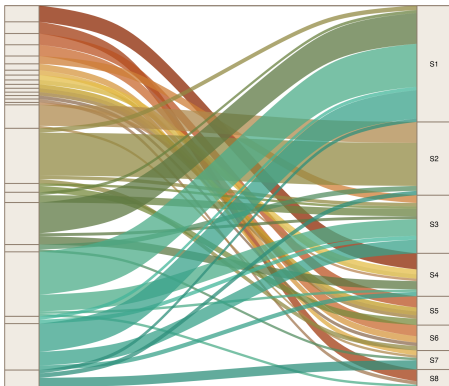
Data and sample construction

Step	2018		2022	
	N	% kept	N	% kept
Total Grade 7 enrolment in LURITS	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
Final analysis sample	532,295	66.8	665,019	76.7

Sample: national ladder, both cohorts. LURITS is the national learner register. Each % kept is of that cohort's total Grade 7 enrolment, so the column reads down as cumulative retention. The final row additionally passes a largest-connected-set restriction, an intervening step this table does not show. The estimation sample is Gauteng and 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. → B1

The transition network mixes densely enough to connect every school



6

secondary schools receive the leavers of the median primary

10

primary schools feed the median Grade 8 class

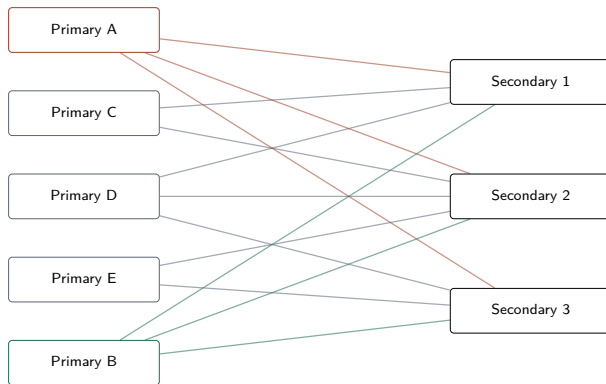
98.8%

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

Roadmap

1. Motivation
2. This paper
3. Related literature
4. Setting and data
- 5. Empirical strategy**
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

Divergence chains every school onto one scale



Leavers scatter; every shared destination is a comparison; comparisons chain into one connected scale, estimated as two-way fixed effects in the AKM tradition.

Anchoring runs both ways: a shared **instrument** at the destination, a shared **scale** at the origin.

Estimating equation

Latent model. Latent learning a is the recorded mark g plus a school-set offset:

$$a_{ij}^7 = g_{ij}^7 + s_j + \varepsilon_{ij}^7$$

$$a_{ijk}^8 = g_{ijk}^8 + d_k + \varepsilon_{ijk}^8$$

$$a_{ijk}^8 = c + \lambda a_{ij}^7 + \nu_{ijk}$$

s_j and d_k are the same object: a school-set price-level offset between true learning and the recorded mark, one applied at the origin and one at the destination. A positive offset means the mark sits below the learning behind it, so $s_j > 0$ is a strict origin.

Estimated by two-way fixed effects:

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

Estimated province by province, on within-province-cohort standardised marks, HC1. Identification comes from pupils who arrive at the *same* destination from *different* feeders. → B17

Mapping estimates onto structural objects

Substituting the three latent equations of the previous frame into one another gives the estimated equation. Its four coefficients are the four rows below, where s_j is the origin's mark-scale offset, d_k the destination's, λ six-month persistence, and ρ the reliability of the Grade 7 mark.

Estimated	Structural target	What it is
β_0	c	six-month growth constant
β_1	λ (plim $\lambda\rho$)	persistence, attenuated by noise
α_j	λs_j	origin mark-scale offset: the calibration object
γ_k	$-d_k$	destination offset: absorbs the destination's own marking regime

- Measurement error in g^7 puts the slope's probability limit at $\lambda\rho$. With $\rho < 1$, $\hat{\alpha}_j$ is attenuated toward zero, so **every magnitude that follows is a floor.** → B2

Identifying assumptions

First, a within-grade anchor. Institutional, and stated in the setting: the whole Grade 8 grade sits the same assessment within a destination, and each feeder grades all its leavers on one scale. Destination side a shared instrument, origin side a shared scale.

Second, a stable short-horizon mapping. Six months, and additivity is testable: quadratic, piecewise-linear, empirical-Bayes and cross-classified specifications recover school effects correlated at 0.987–0.998 with the baseline.

Third, where a pupil goes next does not depend on how much she is about to learn over the six months between the two marks.

- The substantive one, and the only one that can break the wedge.
- School choice as such does not violate it: sorting on the entering mark is absorbed by $\beta_1 g_{ij}^7$, and everything systematic about a destination by γ_k .
- What it rules out is narrower. Two pupils leaving one primary on the same mark must not go to different secondaries *because* one of them is going to learn more over that window.
- The exposure is a feeder-level average, since $\hat{\alpha}_j$ is one number per feeder. A feeder that scatters its leavers averages destination differences away; one that sends them all to a single secondary does not. That is what $\sigma_\gamma \sqrt{H_j}$ measures, and it is a ceiling, not an estimate.

→ B4

The sorting that happens is the sorting the estimator has already removed

Feeder cells differing on	Observed	Placebo	Excess (pp)
Grade 7 baseline mark	22.82%	5.96%	16.9
Gender	5.48%	1.36%	4.1
Over-age	9.76%	3.95%	5.8

Sample: 22,142 feeder-cohort cells; the placebo permutes pupils across destinations within each feeder.

- The placebo re-shuffles pupils across destinations at random within each feeder, holding intake fixed and destroying only the sorting.
- All three margins sit outside their placebo intervals, so none is zero. The point is the ratio: sorting on the conditioned mark runs 2.9 to 4.1 times either demographic margin, and it is the margin $\beta_1 g_{ij}^7$ absorbs before the wedge is formed.
- **This is not the sorting result the deck closes on.** Here the question is whether pupils leaving *one* primary split between secondaries on something the regression has not conditioned on; the market section asks whether an origin's *wedge* predicts the cohort its leavers join.

The design fails only if a destination marks by feeder

Threat	What it would do	What answers it
Destination assignment tracks unobserved growth	Biases the wedge through the pupils a feeder sends rather than through its marking.	Bounded. Worst case $\sigma_{\gamma}\sqrt{H_j}$ (B4); observed sorting runs on the conditioned Grade 7 mark; simulation leaves ranks intact and moves levels (B12).
The Grade 7 to Grade 8 mapping is unstable	Breaks the $\lambda\rho$ reading of β_1 rather than the wedge's ordering.	Bounded. Method correlations at 0.987–0.998. Attenuation from $\rho < 1$ pushes every reported magnitude toward zero, so instability costs magnitude rather than sign.
A destination marks by feeder	Loads straight onto $\hat{\alpha}_j$. No fixed effect removes it, because γ_k absorbs a destination's marking only while that marking is a property of the destination and not of the pupil's origin.	Maintained. The Grade 8 Term 2 mark blends a proctored grade-wide assessment with common continuous assessment tasks, and almost all schools set the same assessments across the grade.

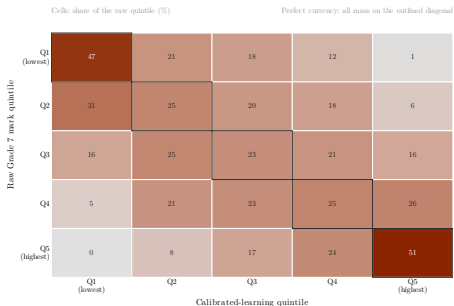
One assumption is maintained rather than bounded, and it is an institutional claim about how a school marks, not a statistical one.

Roadmap

1. Motivation
2. This paper
3. Related literature
4. Setting and data
5. Empirical strategy
- 6. Results: the grading wedge**
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

Two-thirds of schools sit in a different learning quintile

65.8% of schools change learning quintile once the wedge is removed

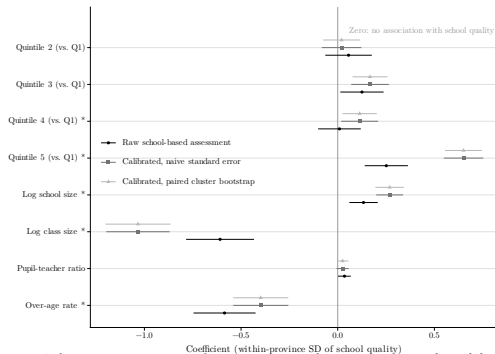


- Only 23% of middle-quintile schools stay put; their mass spreads almost evenly across all five calibrated quintiles.
- The typical school moves 240 ranks; the corners hold, so the movement is mostly local rather than corner to corner.
- **The magnitude behind it.** Grading offsets have a standard deviation of 0.47 on the within-province pupil scale: move a pupil from a school at the 10th percentile of leniency to one at the 90th and her mark shifts by a full within-province SD, with no change in what she knows. → B3

Sample: Gauteng estimation sample, not the national ladder of the data section, and roughly 5 times smaller. Rows may not sum to 100 owing to rounding.

The SES gradient in school quality more than doubles under calibration

One school-level regression, run twice, changing only the left-hand side: the raw mean Grade 7 mark, then the calibrated learning level. Same controls both times.



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

- Quintile 5 schools sit **0.251** SD above Quintile 1 on raw marks, and **0.651** SD above on calibrated learning, on intervals that do not overlap. Log school size doubles the same way, from 0.134 to 0.269.
 - Intervals propagate the first-stage error by paired cluster bootstrap.
 - No-fee schools grade about 0.49 within-province SD more leniently, the only covariate surviving Romano–Wolf.
- Proxying school quality with raw marks credits the richest schools with about half the advantage they hold.**

→ B16

The wedge is signal, not noise

0.88

Split-half correlation in 2022: split each school's pupils at random, calibrate on each half separately, and the two halves rank schools the same way. It rises to 0.94 under Spearman–Brown.

Simulation puts the rank correlation with truth at or above 0.95 at observed densities, and no lower than 0.88 at the worst parameterisations tested. → [B12](#)

0.987

Lowest rank correlation with the baseline across the alternative estimators: empirical Bayes, quadratic, piecewise-linear and cross-classified.

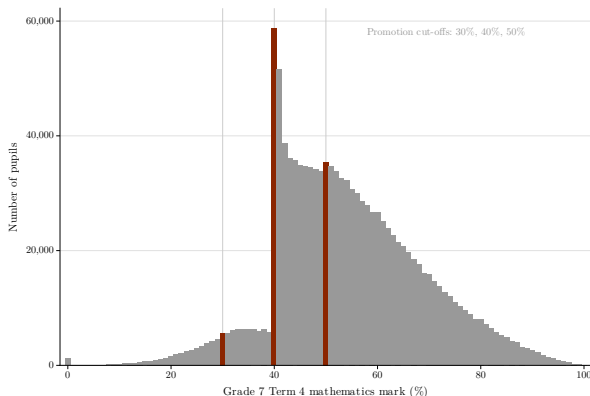
Three reliability measures, and the demanding two are the ones to trust

Each school's wedge is measured with error, and the fewer leavers a school has the larger that error. Reliability is the share of the spread in wedges across schools that is real difference between schools rather than that error. It reads the error off residuals inside the same fit, so it never consults a new pupil or a new cohort, which is why the two rows beneath it are the ones to trust.

Measure	What it asks	2018	2022
Kane–Staiger reliability	Real variation across schools, or sampling noise?	0.970	0.971
Split-half correlation	Does the other random half rank schools the same?	0.884	0.879
Cross-cohort correlation	Does the 2018 ordering hold on the 2022 cohort?	0.602	

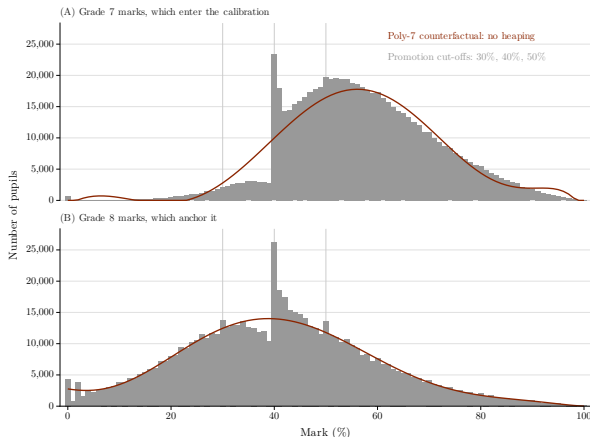
Sample: Gauteng estimation sample. Rows 1 and 2 are estimated per cohort; row 3 is one correlation between the 2018 and the 2022 effects.

Marks heap at the progression thresholds



- Mass piles up at 40, the “adequate achievement” line, and on no other round number to any comparable degree: excess mass there is 126% in 2018 and 147% in 2022, while the 30 and 50 cut-offs, which carry less institutional weight, are near-flat.
- The cliff immediately below 40 is as informative as the spike above it. That asymmetry is what rules out classical measurement error: this error is threshold-specific and correlated with the mark itself.

The heaping does not move the measure



The 2022 cohort. Observed marks in grey, the de-heaped counterfactual in rust over them. Both grades spike at 40.

- Kleven–Waseem de-heaping of the observed data: raw and de-heaped $\hat{\alpha}_j$ correlate at 0.997 or above in all twelve province-cohort cells, and the slope moves by at most 2.1%. → B5 → B14
- Injecting heaping into simulated data at observed severity leaves the rank correlation with truth identical to clean marks.
- **The heaviest-heaping schools grade more strictly across the whole distribution**, which is the opposite of what a manipulation story predicts.

The calibrated level is the durable predictor

2018 predictor ↓ / 2022 target →	Raw Grade 7 mean	Calibrated level
Raw Grade 7 mean	0.441	0.030
Calibrated level	-0.112	0.624

Sample: 1,280 matched Gauteng schools, both cohorts.

- The calibrated level predicts its own future at 0.624 across a four-year, COVID-spanning gap, against 0.441 for raw mean on raw mean.
- The raw 2018 mean predicts the 2022 calibrated level at 0.030, which is almost nothing. For anyone forecasting a school's future standing, the calibrated grade is the stable construct.
- A school's calibrated learning level is mildly negatively related to its later raw mean, at -0.112. That is the bridge identity: to first order $\hat{\alpha}_j \approx \overline{g^8_j} - \beta_1 \overline{g^7_j}$, so a school that grades strictly mechanically shows a lower raw mean.

Same schools, different pupils, four years and a pandemic apart

0.60

Cross-cohort persistence of the wedge, precision-weighted on the matched Gauteng panel ($n = 1,214$).

0.67

Disattenuated correlation between a school's mathematics and English wedges: the trait is not subject-specific. English is excluded from the estimation sample but its wedge is still estimable; the correlation is disattenuated for the lower English reliability.

89.2%

of the variance sits within district, so a district-level marking rubric cannot carry it.

The wedge is a school-level trait, not a cohort accident, not a subject artefact, and not district policy. → B9 → B16

Roadmap

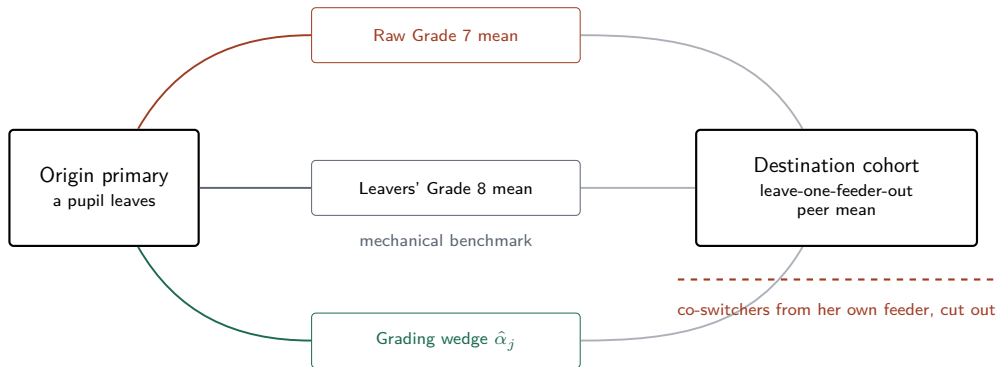
1. Motivation
2. This paper
3. Related literature
4. Setting and data
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
9. Conclusion

Why a market test, and what it settles

- **The problem.** Everything so far says the wedge is large, precise and durable. None of it says the wedge is *learning*, and nothing at primary level is sat by pupils from more than one school, so there is no benchmark to validate it against.
- **The test.** If the calibrated scale is picking up something real, it should show up where families act. Holding a pupil's own Grade 7 mark fixed, does her primary's wedge predict the Grade 8 cohort she ends up in, when its raw mark average does not?
- **What comes next.** The design and its one mechanical threat, the result, then an external check against the only comparable mark in the system, matric.
- **What this cannot do.** It cannot say *why*. A family choosing a distant strong school and a family living beside one leave the same enrolment record.

Families sort as if on the calibrated scale, on a quantity the system has never published and no household could look up.

Research design: three origin signals, one destination outcome



221,119 movers, 1,371 origins, 884 destinations, Gauteng core, both cohorts. Controls include the pupil's own Grade 7 mark; two-way clustered standard errors; all coefficients in within-cohort SD units. Whether households choose or simply live where they live is not identified here. → B7

Enrolment tracks the wedge; the raw mark's edge is own-feeder leakage

Origin signal	Grade 7 intake		Grade 8 achievement	
	Full	LOFO	Full	LOFO
Raw Grade 7 mean	0.310*** (14.4)	0.109*** (4.9)	0.023 (1.6)	0.020 (1.4)
Leavers' Grade 8 mean	0.223*** (7.0)	0.233*** (6.9)	0.601*** (28.7)	0.572*** (25.7)
Grading wedge $\hat{\alpha}_j$	-0.069** (-2.3)	0.102*** (3.6)	0.165*** (9.2)	0.136*** (7.4)

Sample: 221,119 movers, 1,371 origins, 884 destinations; Gauteng core, both cohorts. The leavers' Grade 8 mean is that of the origin's own leavers. Coefficients in within-cohort SD units, t in brackets; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Rows and columns average over different pupil sets: the raw Grade 8 row is 0.6, not 1. Every further de-leaking: → B15

What families see. The origin's raw Grade 7 average predicts **0.020** of the Grade 8 attainment of the cohort her child joins, and is not distinguishable from zero. **What they act on.** The grading wedge predicts **0.136**, on marks written by pupils who never attended her primary: about 7 times the published signal, and the only one of the two that is significant.

External check: the destination's published matric mean

0.144 SD higher published NSC mathematics mean at the destination

That is the gap a 1 SD stricter origin predicts ($t = 7.1$). The NSC is the matric examination, and its school means come from cohorts who left years before the chooser arrived, so no own-feeder machinery can reach this outcome.

Origin signal	Coefficient	SE	t
Grading wedge $\hat{\alpha}_j$	0.144	0.020	7.1
Raw Grade 7 mean	0.049	0.015	3.2
Leavers' Grade 8 mean	0.224	0.027	8.3

Sample: pooled single-predictor fits; 209,666 movers, 756 destinations. Outcome: the destination's published NSC mathematics mean, standardised within cohort.

The wedge outruns the raw Grade 7 mark families are shown by about 3 to one, on intervals that do not overlap. → B8

Roadmap

1. Motivation
2. This paper
3. Related literature
4. Setting and data
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
- 8. Robustness**
9. Conclusion

Robustness checks: where each one lives

Threat	Check	Where
Pupils sort to destinations on unobservables	Herfindahl bound; permutation placebo; Monte Carlo under sorting	B4, B12
Limited-mobility bias, thin networks	Algebraic connectivity against the Jochmans–Weidner benchmark	B11
Mark manipulation at thresholds	Kleven–Waseem de-heaping, both grades; downstream gradients	B5, B14
Attrition at record linkage	Cut-off sweep across truncation percentiles; IPW	B1
Gauteng only	Six-province panel, pooled and by province	B6
Households relocate rather than choose	Cross-ward and cross-municipality switchers; distance interaction	B7
The wedge is a generated regressor	Paired cluster bootstrap, wedge re-estimated per replicate	main deck
The Grade 7 to Grade 8 mapping is non-linear	Quadratic, piecewise-linear, empirical-Bayes, cross-classified	B17
The sorting result rests on one de-leaking cut	Cross-cohort substitution, own-feeder cut, destination leniency netted out	B15

Roadmap

1. Motivation
2. This paper
3. Related literature
4. Setting and data
5. Empirical strategy
6. Results: the grading wedge
7. Results: does the market price the wedge?
8. Robustness
- 9. Conclusion**

What a calibrated mark is for, and what it is not for

Supports

- Comparing schools on one scale, across the whole system.
- Identifying strict graders whose pupils are under-recognised.
- Forward prediction, on the durable ordering.
- A zero-marginal-cost partial substitute for a discontinued national assessment, built from records the system already holds.

Does not support

- Accountability or effectiveness attribution: the wedge is a composite.
- Absolute learning levels: the scale is relative and cohort-specific.
- Claims about how households deliberate.

The design travels wherever a transition mixes graduates across assessment regimes and the receiving assessment is the same for every pupil in a destination's grade. No common test across destinations is needed. That covers European tracked lower-secondary transitions and LMIC systems where school-set marks are the only measure at scale.

Conclusion

Schools grade in **local currencies**. The primary-to-secondary transition puts them on a **common currency** and reveals each school's exchange rate. Households already sort on that exchange rate, not on the grades it converts.

Backup index

Identification and sorting

- B4 Herfindahl bound on destination sorting
- B11 Network density and limited-mobility bias
- B12 Monte Carlo: ranks robust, levels fragile
- B17 The latent model in full

Sample and measurement

- B1 Sample funnel and attrition cut-off sweep
- B2 The $\lambda\rho$ decomposition
- B5 De-heaping the observed data
- B14 What heaping does downstream

The sorting test's own robustness

- B6 Joint specification and six-province panel
- B7 The geography bound
- B15 The sorting result survives every de-leaking

Downstream and comparative

- B3 The full two-way fixed-effects table
- B8 The NSC external criterion
- B9 Spatial clustering and mean reversion
- B10 How this differs from causal value-added
- B13 Lenient graders are not simply weak schools
- B16 What the wedge is made of

B4 · The Herfindahl bound on destination sorting

- The estimator's exposure to unobserved growth is at the school level. A feeder's bias depends on $\bar{\gamma}_j = \sum_k \pi_{jk} \gamma_k$, the average destination effect its leavers face, so the worst case is $\sigma_\gamma \sqrt{H_j}$, with H_j the feeder's Herfindahl concentration across destinations.
- A feeder sending every leaver to one secondary is fully exposed to that school's effect. One that spreads its leavers wide averages destination differences away.
- Observable-characteristics placebo: each pupil's SES-predicted Grade 8 mark, averaged to their origin school and regressed on the wedge, gives a slope of 0.515 and an R^2 of 0.421. Near zero would reassure if the estimand were a causal teacher effect. It is not: this estimand is a composite that includes student selection, so a positive slope is expected, and it is the level that stays exposed, not the ranks.
- **Answers “this is Rothstein”** (Rothstein 2010). The bound is the worst case; the permutation test and B12's simulation are what say the realised bias is small.

← Identifying assumptions

B11 · 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 $\lambda_2 = 0.2623$, about 67 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 Andrews and Kline-style corrections that bite in sparse labour networks are second-order here.

[← Related literature](#)

B12 · Monte Carlo: ranks robust, levels fragile

Threat	Scenario	Rank correlation			Level accuracy	Quintile stability	Verdict
		Base	Worst	(marks)			
Measurement error (heaping)	Heaping index 1–3	0.968	0.968	<0.001		74.7%	Negligible
Limited mobility (sparsity)	2–11 destinations, 15–80 students	0.968	0.876	<0.001		74.6%	Negligible at observed density
Endogenous mobility (Andrews/KSS)	$\psi = 0$ to 2, observed density	0.963	0.897	RMSE $\psi=2$	2.51 at	58.5%	Ranks stable; levels degrade
Non-random attrition	Random to extreme selection	0.960	0.960	<0.001		72.0%	Negligible
Sorting + peer effects	$\rho=0-1$, $\lambda=0-0.15$	0.953	0.985	<0.001 (RMSE 5.84 at $\rho=1$)		79.4%	Ranks stable; levels degrade

Sample: Level accuracy is the mean bias where that is the binding statistic, and the worst-case RMSE with its parameter value where the levels move. Rank correlations are against the simulated truth.

- Sorting stretches the ruler and endogenous mobility shrinks it. Both leak quality between feeder and destination effects, and neither reorders schools.
- The sorting row's worst case reads *higher* than its baseline, because stretching moves true school differences apart while estimation noise stays fixed, so neighbouring reversals become rarer.
- Bounding: the empirical feeder-to-destination mixing correlation is 0.021 and would have to reach 0.048, about 2.3 times stronger, before the bias reached the 5% threshold.

← Signal, not noise

B17 · The latent model in full

$$a_{ij}^7 = g_{ij}^7 + s_j + \varepsilon_{ij}^7$$

$$a_{ijk}^8 = g_{ijk}^8 + d_k + \varepsilon_{ijk}^8$$

$$a_{ijk}^8 = c + \lambda a_{ij}^7 + \nu_{ijk}$$

Substituting and relabelling the structural composites gives

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

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**:

$$g_{ij}^7 = a_{ij}^7 - s_j - \varepsilon_{ij}^7, \quad \varepsilon_{ij}^7 \perp a_{ij}^7$$

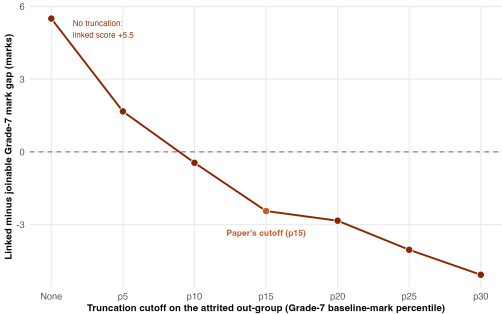
so ε^7 sits inside the regressor and
 $\rho = \text{Var}(\tilde{a}^7)/\text{Var}(\tilde{g}^7)$ is the mark's reliability.

- 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 $\hat{\alpha}_j \approx \bar{g}_j^8 - \beta_1 \bar{g}_j^7$, 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.

← Estimating equation

B1 · Sample funnel and the attrition cut-off sweep

Step	2018	2022
Total Grade 7 enrolment in LURITS	796,928	866,679
With valid Grade 7 Term 4 maths mark	708,659 (88.9% of total)	804,654 (92.8% of total)
Successfully linked to a Grade 8 record	544,233 (76.8% of valid Gr7)	674,557 (83.8% of valid Gr7)
With valid Grade 8 Term 2 maths mark	544,233 (100.0% of linked)	674,557 (100.0% of linked)
In the largest connected set	542,911 (99.8% of valid Gr8)	673,896 (99.9% of valid Gr8)
Final analysis sample	532,295 (98.0% of connected)	665,019 (98.7% of connected)



- The gap's **sign** is a choice of truncation. It runs from +5.5 marks with no truncation, through zero at about the ninth percentile, to roughly -4.5 at the thirtieth. Quoting any single number without its cut-off would be meaningless.

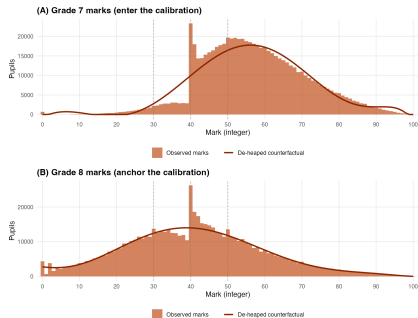
← Data and sample

B2 · 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 $\rho \approx 0.63$ to 0.71 and persistence $\lambda \approx 0.85$ to 0.96 , 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.

← Mapping estimates

B5 · De-heaping the observed data



Cohort	Mark	Observed	Counterfactual	Excess	Index
2018	30	3,411	3,109	9.7%	1.11
2018	40	35,380	15,657	126.0%	1.93
2018	50	15,708	15,014	4.6%	1.01
2022	30	2,187	2,114	3.5%	1.06
2022	40	23,337	9,467	146.5%	2.24
2022	50	19,681	18,662	5.5%	1.05

Sample: Index is the count at that mark over the mean of the counts at the two adjacent marks. Excess compares the same count with the smooth counterfactual, which is fitted off the non-multiples of five, so the two columns use different denominators and do not agree.

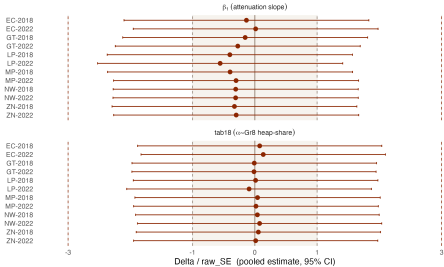
- Panel A is the Grade 7 mark that enters the calibration, panel B the Grade 8 mark that anchors it. **Both sides of the estimating equation heap**, which is the form of the objection that actually threatens the wedge.
- Panel B's counterfactual sits well below the observed mass between 40 and 55, so the redistribution is substantial even though it does not move the measure.

← De-heaping

B14 · What heaping does downstream

Structural de-heaping does not move β , or tab18 in any cell

Rubin-pooled Δ expressed in units of the raw standard error
(12 province-by-cohort cells, $m = 25$ imputations).



Source: stage7_rubin_pooled_deltas.csv

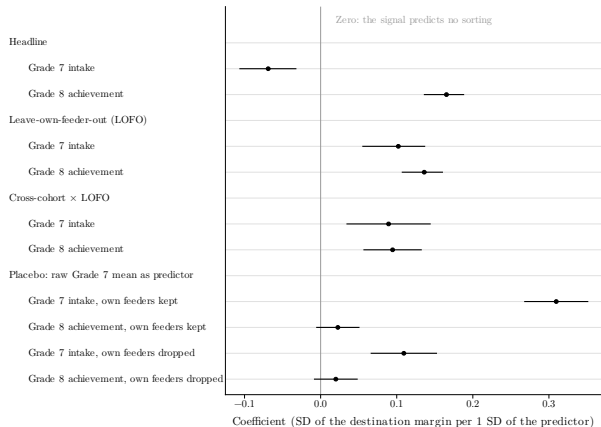
Top: the attenuation slope. Bottom: the coefficient on a school's Grade 8 heap share. Across all 12 province-cohort cells, neither moves by more than 0.56 of a raw standard error.

Covariate	(A) Raw Gr7 mean	(B) Baseline calibrated	(C) Multi-term calibrated	(D) De-heaped calibrated
Quintile 2	-0.040 (0.700)	-0.444 (0.657)	-0.076 (0.641)	-0.439 (0.654)
Quintile 3	0.294 (0.634)	0.581 (0.595)	0.989 (0.581)	0.583 (0.592)
Quintile 4	-0.443 (0.621)	1.920 (0.582)	2.090 (0.568)	1.926 (0.579)
Quintile 5	0.491 (0.672)	6.432 (0.630)	6.706 (0.616)	6.454 (0.628)
Log enrolment	-0.115 (0.447)	2.359 (0.419)	2.151 (0.409)	2.338 (0.417)
Pupil-teacher ratio	-0.140 (0.043)	-0.460 (0.041)	-0.464 (0.040)	-0.460 (0.040)
Over-age rate	-25.190 (2.155)	5.943 (2.021)	6.241 (1.973)	5.743 (2.012)
R ²	0.146	0.334	0.346	0.337
RMSE	6.569	6.159	6.015	6.133

- Column A is the raw Grade 7 mean, B the baseline calibrated measure, C a multi-term calibration and D the calibrated measure rebuilt on de-heaped marks. Standard errors in brackets.
- Columns B and D are near-identical throughout: de-heaping the input does not move the gradients the calibrated measure recovers.
- The over-age row is where raw and calibrated disagree most, and it changes sign.

← De-heaping

B15 · 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.

Green here marks the most severe specification, not “calibrated”. The four colours are specification families.

← Enrolment tracks the wedge

B6 · Joint specification and the six-province panel

Province	n	Full	LOFO	<i>t</i>
EC	162,359	0.068	0.012	0.7
GT	220,995	0.166	0.135	7.2
LP	182,361	0.107	0.043	2.5
MP	106,965	0.143	0.091	4.3
NW	91,125	0.117	0.047	2.4
ZN	254,524	0.069	0.013	0.8
POOLED6	1,018,329	0.116	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.

← Backup index

B7 · The geography bound

- Among cross-ward switchers the achievement margin holds at 0.143 ($t = 7.7$); among cross-municipality switchers at 0.081 ($t = 3.0$). 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.

← Research design

B3 · The full two-way fixed-effects table

Metric	2018	2022
Pupils	106,514	135,852
Grade 7 schools	1,424	1,479
Grade 8 grades	901	839
β_1 (Grade 7 mark coefficient)	0.601	0.592
$SE(\beta_1)$	0.003	0.002
Adjusted R^2	0.678	0.609
Within R^2	0.404	0.398
SD of school effects	0.471	0.468
Median SE of school effects	0.067	0.067
% of schools with SE below 0.5 SD	100.0%	100.0%
% of schools significantly $\neq 0$ ($p < 0.05$)	70.4%	74.8%
Signal-to-noise ratio (SD / median SE)	7.00	6.99

Sample: Gauteng estimation sample, both cohorts. The column is row-heterogeneous, so decimals are uniform within a row rather than down the column.

The school counts elsewhere in the deck are different restrictions on the same data, not disagreements. The 1,424 and 1,479 above are the Grade 7 schools carrying an estimated effect in this Gauteng sample, one count per cohort. The 1,371 origins on the research-design frame are the mover panel pooled over both cohorts, which also needs a linked destination, complete controls and a non-missing destination peer mark. The 1,280 behind the calibrated-level prediction are the schools matched across both cohorts with a raw baseline mark in each. The 1,214 behind cross-cohort persistence are the smaller set matched for the cross-cohort covariance.

← Reclassification

B8 · The NSC external criterion

- This correlation sets each secondary's own Grade 8 mean against its own matric record, so it tests the destination means rather than the origin wedge, and a single-year matric mean is a noisy criterion for either. The pupil-level coefficient on the main frame is the statistic that bears on the claim: it asks whether an *origin's* wedge predicts the published mean of the destination its leavers join.
- Rank correlation of the destination's published NSC mathematics mean with the origin signal: raw 0.232 in 2018 and 0.261 in 2022, against calibrated 0.232 and 0.267.
- The criterion is noisy by construction. A school's matric mean reflects its admissions and its attrition as much as its teaching, and it is measured on cohorts who entered five years before.
- That noise cuts one way only: it attenuates the criterion correlation.

← External check

B9 · Spatial clustering and mean reversion

- Moran's I at a 5 km threshold is 0.197 in 2018, so lenient schools cluster geographically to a modest degree.
- Cross-cohort convergence is -0.513, which is an **upper bound** on true convergence rather than an estimate of it: estimation error in the first cohort's wedge mechanically produces mean reversion in the second.
- Both are descriptive, and neither identifies a mechanism.

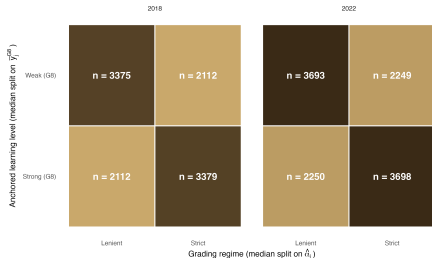
← Durability

B10 · How this differs from causal value-added

- Liu et al. (2025), and the value-added literature generally, start from a test every pupil has already sat, and use the receiving institution to attribute later outcomes back to the sending school.
- This design has no such test. It uses the receiving institution to *build* the scale, and stops there: it removes only the *scale* on which a school reports marks, and leaves selection inside the measure.
- That is why $\hat{\alpha}_j$ is a composite and why the paper claims no effectiveness attribution. The two measures answer different questions and should not be substituted for one another.
- The object here is a price, not a production-function coefficient.

[← Backup index](#)

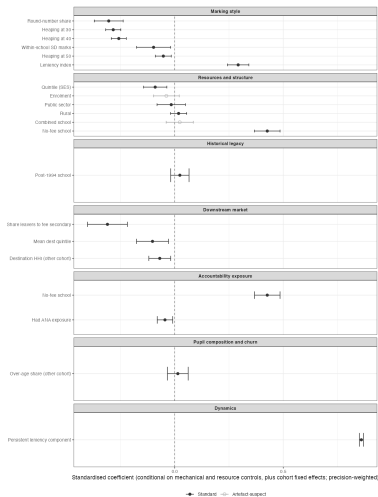
B13 · Lenient graders are not simply weak schools



- A 2 x 2 mosaic of **median splits**, on $\hat{\alpha}_j$ and on the raw Grade 8 cohort mean, taken within each province-cohort cell. Cell labels are school counts. It contains no quintiles.
- The sample is the national one across both cohorts, not the Gauteng estimation sample, which is why the counts sum to far more schools than the main table reports.
- The wedge correlates with the raw Grade 7 mean at -0.58 , partly by the bridge identity.
- Strict grading and the strong-learning half do associate: the odds ratio is about 2.6 in 2018 and 2.7 in 2022. The claim is not independence.
- **There are lenient schools at every raw-mark level**, so calibration corrects a pricing error rather than relabelling weak schools.

← Backup index

B16 · What the wedge is made of



Family	Share of joint R^2
F6 Downstream market	38.9%
F3 Resources and structure	28.6%
F7 Accountability exposure	28.1%
F8 Pupil composition and churn	2.6%
F4 Historical legacy	1.8%

- Three families carry 96% of the joint R^2 of 37.4%, and downstream market leads without commanding.
- No-fee schools grade about 0.49 within-province SD more leniently, the only covariate surviving Romano-Wolf.
- Each 1 SD of destination-market concentration associates with 0.031 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.

References I

- Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review* 110 (5): 1502–39.
<https://doi.org/10.1257/aer.20172040>.
- Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333. <https://doi.org/10.1111/1468-0262.00020>.
- Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." *American Economic Review* 107 (6): 1535–63.
<https://doi.org/10.1257/aer.20140774>.
- Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics." *American Economic Journal: Applied Economics* 3 (3): 29–54.
<https://doi.org/10.1257/app.3.3.29>.

References II

- Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward. 2008. “High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (3): 673–97. <https://doi.org/10.1111/j.1467-985X.2007.00533.x>.
- Araujo, M. Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady. 2016. “Teacher Quality and Learning Outcomes in Kindergarten.” *The Quarterly Journal of Economics* 131 (3): 1415–53. <https://doi.org/10.1093/qje/qjw016>.
- Bau, Natalie, and Jishnu Das. 2020. “Teacher Value Added in a Low-Income Country.” *American Economic Journal: Economic Policy* 12 (1): 62–96. <https://doi.org/10.1257/pol.20170243>.
- Bergbauer, Annika B., Eric A. Hanushek, and Ludger Woessmann. 2024. “Testing.” *Journal of Human Resources* 59 (2): 349–88. <https://doi.org/10.3368/jhr.0520-10886R1>.
- Bertoni, Marco, Giorgio Brunello, and Lorenzo Rocco. 2013. “When the Cat Is Near, the Mice Won’t Play: The Effect of External Examiners in Italian Schools.” *Journal of Public Economics* 104: 65–77. <https://doi.org/10.1016/j.jpubeco.2013.04.010>.

References III

- Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114 (2): 577–99. <https://doi.org/10.1162/003355399556070>.
- Bold, Tessa, Deon Filmer, Gayle Martin, et al. 2017. "Enrollment Without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa." *Journal of Economic Perspectives* 31 (4): 185–204. <https://doi.org/10.1257/jep.31.4.185>.
- Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128 (3): 967–1015. <https://doi.org/10.1093/qje/qjt006>.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–632. <https://doi.org/10.1257/aer.104.9.2593>.
- Dee, Thomas S., Will Dobbie, Brian A. Jacob, and Jonah Rockoff. 2019. "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations." *American Economic Journal: Applied Economics* 11 (3): 382–423. <https://doi.org/10.1257/app.20170520>.

References IV

- Diamond, Rebecca, and Petra Persson. 2016. *The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests*. NBER Working Paper No. 22207. National Bureau of Economic Research. <https://doi.org/10.3386/w22207>.
- Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review* 94 (3): 591–604. <https://doi.org/10.1257/0002828041464489>.
- Gustafsson, Martin. 2019. *Costs and Benefits of Different Approaches to Measuring the Learning Proficiency of Students (SDG Indicator 4.1.1)*. UIS Information Paper No. 53. UNESCO Institute for Statistics. [https://www.uis.unesco.org/sites/default/files/medias/fichiers/2025/08/GAML6-REF-7-Cost-and-benefit-approaches-to-4.1.1%20\(1\).pdf](https://www.uis.unesco.org/sites/default/files/medias/fichiers/2025/08/GAML6-REF-7-Cost-and-benefit-approaches-to-4.1.1%20(1).pdf).
- Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *The Quarterly Journal of Economics* 123 (4): 1373–414. <https://doi.org/10.1162/qjec.2008.123.4.1373>.

References V

- Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118 (3): 843–77. <https://doi.org/10.1162/00335530360698441>.
- Jochmans, Koen, and Martin Weidner. 2019. "Fixed-Effect Regressions on Network Data." *Econometrica* 87 (5): 1543–60. <https://doi.org/10.3982/ECTA14605>.
- Kane, Thomas J., and Douglas O. Staiger. 2008. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation*. NBER Working Paper No. 14607. National Bureau of Economic Research. <https://doi.org/10.3386/w14607>.
- Kline, Patrick, Raffaele Saggio, and Mikkel Sølvesten. 2020. "Leave-Out Estimation of Variance Components." *Econometrica* 88 (5): 1859–98. <https://doi.org/10.3982/ECTA16410>.
- Liu, Jing, Seth Gershenson, and Max Anthenelli. 2025. *Ready or Not? A New Way to Measure Elementary and Middle School Quality*. Report. Thomas B. Fordham Institute.

References VI

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review* 105 (11): 3471–88. <https://doi.org/10.1257/aer.20130332>.

MacLeod, W. Bentley, and Miguel Urquiola. 2019. "Is Education Consumption or Investment? Implications for School Competition." *Annual Review of Economics* 11: 563–89. <https://doi.org/10.1146/annurev-economics-080218-030402>.

Mansfield, Richard K. 2015. "Teacher Quality and Student Inequality." *Journal of Labor Economics* 33 (3): 751–88. <https://doi.org/10.1086/679683>.

Rossiter, Jack, and Maimouna Konate. 2022. *Studying School Exams: A New Database*. CGD Note. Center for Global Development. <https://www.cgdev.org/publication/studying-school-exams-new-database>.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics* 125 (1): 175–214. <https://doi.org/10.1162/qjec.2010.125.1.175>.

References VII

Singh, Abhijeet. 2015. "Private School Effects in Urban and Rural India: Panel Estimates at Primary and Secondary School Ages." *Journal of Development Economics* 113: 16–32.
<https://doi.org/10.1016/j.jdeveco.2014.10.004>.

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. World Bank.
<https://doi.org/10.1596/978-1-4648-1096-1>.