

The Systemic Obsolescence of Global Education Systems: A Root Cause Analysis of Institutional Inertia and Structural Misalignment

A systems account of why school reforms rarely change how schools work: five mutually reinforcing mechanisms, the evidence for and against each, an India case study, and testable propositions for institutional redesign.

Saket Poswal

Author

Sep 20, 2025 • Updated: Sep 29, 2026

ABSTRACT

Education systems have expanded enrolment dramatically, yet many students leave school without the skills schooling is meant to provide, and reforms that target curricula, technology or funding repeatedly leave the core practices of schooling unchanged. This paper offers a systems account of that stability. Drawing on history, economics and sociology, it identifies five mechanisms that reinforce one another: (1) the persistence of an institutional "grammar of schooling" (age-graded classes, fixed timetables, uniform curricula); (2) credentialism, in which schooling serves partly as a sorting and signalling device; (3) assessment-curriculum feedback, in which high-stakes measures narrow instruction; (4) multitask incentive problems, in which rewards attach to what is measurable rather than to learning; and (5) a scale-personalisation trade-off embedded in cohort-based instruction. For each mechanism the paper sets out supporting evidence and, unlike polemical accounts, the most important counter-evidence, including findings that school spending and accountability can improve outcomes. An India case study, using ASER data and randomised evaluations of Teaching at the Right Level and the Mindspark adaptive-learning programme, shows both how the mechanisms operate and how they can be weakened. The paper concludes with six testable propositions and design principles for reforms that change the structure, not only the content, of schooling.

Author

Saket Poswal · [ORCID 0009-0009-8574-9953](#) · [Google Scholar](#)

Type

Analysis (working paper, not peer reviewed)

Published	20 September 2025
Last revised	29 September 2026
Method	Conceptual systems analysis and critical narrative review of historical, economic and sociological evidence; India case study using secondary data and published randomised evaluations
License	CC BY 4.0
DOI	10.5281/zenodo.17812184
Keywords	grammar of schooling · education reform · institutional inertia · credentialism · signalling · high-stakes testing · Campbell's law · multitask principal-agent · personalised learning · Teaching at the Right Level · India NEP 2020 · learning crisis

Working paper · Version 2.0 (29 September 2026)

Revision note. Version 2.0 substantially revises the September 2025 paper. The five-mechanism argument is retained, but each mechanism is now tested against counter-evidence the original omitted (on school spending, test-based accountability and tutoring effect sizes). The historical account of the “Prussian model” has been corrected against historical scholarship; the India case study now rests on cited data and randomised evaluations; and the conclusion is restated as testable propositions. A nonexistent reference was replaced in an earlier revision.

1. Introduction

Over the past half-century, school enrolment has expanded faster than at any time in history. Yet a large share of children in school are not learning the basics: the World Bank (2018) describes a global “learning crisis” in which schooling and learning have come apart, a gap Pritchett (2013) summarised as “schooling ain’t learning.” In rural India in 2024, only 23.4% of children in Class III could read a Class II-level text (ASER Centre, 2025).

The problem is not a shortage of ideas about how people learn. Evidence has long favoured instruction matched to the learner’s current level, mastery-based progression, and conditions that support intrinsic motivation (Bloom, 1984; Guskey, 2010; Deci & Ryan, 2000). The puzzle is why schools so rarely organise themselves around this evidence, and why reform after reform changes what is taught while leaving *how schooling is organised* largely intact. Historians of education call this stable core the “grammar of schooling” (Tyack & Tobin, 1994; Tyack & Cuban, 1995).

This paper offers a systems account of that stability. Its central claim is that education systems are held in place by several mechanisms that reinforce one another, so that a reform acting on one is absorbed by the others. The paper's contributions are:

1. Five mechanisms of institutional stability, drawn from history, economics and sociology, with the causal pathway of each made explicit (Section 3).
2. A balanced evidence review that tests each mechanism against the strongest counter-evidence, instead of assembling only supporting sources (Section 4).
3. An India case study using ASER data and randomised evaluations, showing both the mechanisms at work and interventions that weaken them (Section 5).
4. Six testable propositions and a set of design principles for reforms that target structure rather than content (Sections 6–7).

The account is conceptual. It does not present new data, and the mechanisms are offered as hypotheses consistent with the evidence rather than as a demonstrated root-cause model.

2. Historical Context: Where the Grammar of Schooling Came From

2.1 The “Prussian factory model”, corrected

A popular narrative holds that modern schools were designed in Prussia to produce obedient factory workers and were copied worldwide (e.g., Gatto, 1992). Historians give a more complicated account. Compulsory schooling in Prussia and Austria developed in the eighteenth century under absolutist rulers, before industrialisation, as part of a broader effort to strengthen state authority and religious and civic discipline, and it met substantial local resistance (Melton, 1988). The age-graded, bureaucratically administered school that became the international norm was consolidated later, particularly in nineteenth- and early twentieth-century American cities, where reformers sought efficiency and uniformity for rapidly growing, diverse populations (Tyack, 1974).

The correction matters for the argument. The durable features of schooling (age-graded classes, fixed timetables, subject periods, uniform curricula and grades) were not the output of a single design intended to produce compliant labour. They accumulated as administrative solutions to the problem of educating very large numbers of children, and they then became taken-for-granted rules of what a “real school” is (Tyack & Tobin, 1994).

2.2 Sorting alongside teaching

Mass schooling also acquired a second function: sorting students into different futures. Rosenbaum (1976) documented how tracking channelled students into unequal trajectories, and Bowles and Gintis (1976) argued that schools' social relations corresponded to those of the workplace. One need not accept every part of these arguments to recognise that schools both *teach* and *select*, and that the two functions

can pull in different directions: teaching is best served by helping everyone succeed, while selection requires differentiation.

3. Five Mechanisms of Institutional Stability

3.1 M1: The grammar of schooling

Tyack and Tobin (1994) examined reforms, including the Dalton Plan and flexible-scheduling high schools of the 1960s, that tried to change the basic organisation of schooling, and found that most were abandoned or absorbed. The grammar persists because it is embedded in law, in teacher training, in buildings and timetables, and in the public's cultural understanding of what school should look like; reforms that violate it look illegitimate to parents and teachers even when they work. Tyack and Cuban (1995) describe a century of reform as "tinkering": schools adapt reforms to themselves more often than reforms change schools.

Pathway: reform introduced → conflicts with established organisational routines and public expectations → adapted or abandoned → grammar reinforced.

3.2 M2: Credentialism and signalling

Collins (1979) argued that rising educational requirements for jobs reflect competition for status as much as the skills jobs demand; Brown (2001) traced how credential requirements became institutionalised in labour markets. Economic theory distinguishes two ways schooling raises earnings: by building productive skills (human capital; Becker, 1964) and by *signalling* pre-existing ability and persistence to employers (Spence, 1973). Caplan (2018) argues that signalling accounts for a large share of the return to education. If signalling is important, institutions face incentives to preserve the *scarcity and rank-ordering* of credentials, not only the learning behind them.

Pathway: employers screen on credentials → students and families demand credentials → institutions compete on selectivity and rank → learning becomes instrumental to the credential.

3.3 M3: Assessment–curriculum feedback

When high stakes attach to test results, instruction tends to narrow toward what is tested. A metasynthesis of qualitative studies found that high-stakes testing narrowed curricular content, fragmented knowledge into test-related pieces and increased teacher-centred pedagogy (Au, 2007). Scores can rise faster than the learning they are meant to indicate, the phenomenon of score inflation described by Koretz (2017), and in extreme cases incentives produce outright manipulation: Jacob and Levitt (2003) estimated that serious teacher cheating occurred in a small but meaningful share of Chicago classrooms under high-stakes accountability. These are instances of Campbell's law: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it

will be to distort and corrupt the social processes it is intended to monitor” (Campbell, 1979; see also Muller, 2018; Nichols & Berliner, 2007).

Pathway: high-stakes measure introduced → instruction reallocated toward measured content → measured scores rise faster than learning → measure loses validity → pressure for more measurement.

3.4 M4: Multitask incentive problems

Holmström and Milgrom’s (1991) multitask principal–agent model explains why strong incentives on measured performance can be harmful. When an agent performs several tasks but only some are measured, rewarding the measured tasks draws effort away from the unmeasured ones, so the optimal contract may use weak incentives. Teaching is the canonical case: test scores capture some goals of education, while others, such as curiosity, reasoning, collaboration and long-run understanding, are hard to measure. The same logic applies at each level of the system. Teachers, administrators and politicians face rewards tied to visible, short-run indicators (scores, enrolment, buildings, announcements) more than to long-run learning.

Pathway: only some outcomes measurable → incentives attach to measured outcomes → effort shifts away from unmeasured learning goals → apparent success on measures masks stagnation.

3.5 M5: The scale–personalisation trade-off

Bloom (1984) reported that one-to-one tutoring combined with mastery learning raised achievement by about two standard deviations over conventional instruction, and posed the “2 sigma problem” of achieving similar results at group scale. Later syntheses find a smaller but still large effect: VanLehn (2011) estimated the effect of human tutoring at about $d = 0.79$ and of step-based intelligent tutoring systems at about 0.76. Cohort-based instruction, one teacher, one pace, one curriculum for thirty to sixty students, cannot easily provide this. The consequence is sharpest where learning levels within a class vary widely, as they do in many low- and middle-income countries: a curriculum pitched at grade level leaves behind the students who most need instruction at their own level (Pritchett, 2013; Banerjee et al., 2017).

Pathway: instruction organised by age cohort → curriculum paced to grade level → students below grade level fall further behind → learning variance widens → grade-level teaching becomes less appropriate for more students.

3.6 How the mechanisms reinforce one another

The mechanisms are not independent. Credentialism (M2) raises the stakes of assessment (M3); high-stakes assessment defines what incentives reward (M4); incentives tied to cohort-level results entrench cohort-based organisation (M5); and all four are protected by the legitimacy of the familiar grammar of schooling (M1). A reform aimed at one mechanism, for example a new competency-based curriculum, is absorbed by the others: it is assessed with the old examinations, delivered within the old timetable, and judged by the old credentials. This is the systems-level sense in which reforms “address symptoms”.

4. Testing the Account Against Counter-Evidence

A systems account is only credible if it survives the strongest evidence against it. The original version of this paper omitted several important findings; they are addressed here.

4.1 Does money matter?

Hanushek (2003) reviewed evidence that increases in conventional inputs, such as spending and class-size reductions, had weak or inconsistent relationships with achievement. The original version of this paper used that finding to argue that resources are largely wasted. Later quasi-experimental evidence is more positive. Using court-ordered school finance reforms in the United States, Jackson, Johnson and Persico (2016) found that increased per-pupil spending raised educational attainment and adult earnings, with larger effects for children from low-income families. Claims that administrative staffing has grown faster than teaching staff (Scafidi, 2012) are descriptive and have been disputed on interpretation.

Implication for the account: resources are not irrelevant; *how* they are used within existing structures matters. The account predicts that spending raises outcomes most when it changes what happens between teachers and students, and least when it expands structures that leave instruction unchanged.

4.2 Does test-based accountability always harm?

The original version cited Ravitch (2010) for the claim that No Child Left Behind reduced quality. The causal evidence is more mixed. Dee and Jacob (2011) found that NCLB produced statistically significant gains in fourth-grade mathematics, especially for lower-achieving students, but no gain in reading. At the same time, evidence of curriculum narrowing, score inflation and manipulation (Au, 2007; Koretz, 2017; Jacob & Levitt, 2003) indicates real costs.

Implication for the account: M3 and M4 describe distortions *alongside* genuine effects. Accountability can raise measured skills that are well aligned with tests while degrading unmeasured ones, which is exactly what the multitask model predicts.

4.3 Is technology the answer, or part of the problem?

Cuban (2001) documented how computers were widely purchased but used in ways that left classroom practice largely unchanged, an instance of M1 absorbing a reform. But not all educational technology is alike. In a randomised evaluation in Delhi, the adaptive-learning programme Mindspark, which pitches instruction at each student's measured level, raised mathematics and Hindi scores by 0.37 and 0.23 standard deviations after about four and a half months (Muralidharan et al., 2019). The difference between these findings is itself evidence for the account: technology that reproduces cohort instruction on a screen changes little, while technology that breaks the link between age cohort and instructional level (M5) can change a great deal.

4.4 International comparisons

Finland's high performance in early PISA cycles has been attributed to high teacher autonomy, limited standardised testing and low competitive pressure (Sahlberg, 2015). Such cross-national comparisons are suggestive but weak causal evidence, since systems differ in many ways at once, and Finland's own results have declined in later PISA cycles. In England, the expansion of academies has not been shown to improve outcomes once intake differences are accounted for, and has been linked to patterns of socio-economic segregation (Gorard, 2014). These cases illustrate the mechanisms but cannot test them.

5. Case Study: India

5.1 The learning gap

India achieved near-universal primary enrolment, but learning has lagged behind. ASER surveys of rural households have documented this for two decades. In 2024, 23.4% of Class III children in rural India could read a Class II-level text, recovering from a pandemic-era decline in government schools (from 16.3% in 2022; ASER Centre, 2025). School closures during COVID-19 produced learning losses internationally, averaging about 0.14 standard deviations and larger for disadvantaged children (Betthäuser et al., 2023).

5.2 The mechanisms in operation

- M5 (curriculum pace vs. learning level). When many children in a class are several grade levels behind, instruction paced to the textbook leaves them behind. Pritchett (2013) and Banerjee et al. (2017) identify over-ambitious, grade-paced curricula as a central cause of low learning in India.
- M2 and M3 (credentials and high-stakes examinations). Board examinations and competitive entrance examinations for higher education and public employment carry very high stakes, which pulls instruction toward examination preparation and supports a large private coaching sector.
- M1 (institutional grammar). The National Education Policy 2020 calls for foundational literacy and numeracy, competency-based learning and reduced examination pressure (Ministry of Human Resource Development, 2020). The account predicts that these aims will be absorbed unless examination structures, teacher incentives and classroom organisation change with them. That is a testable prediction for the coming years.

5.3 Interventions that weaken the mechanisms

India also offers some of the strongest evidence that the mechanisms can be weakened:

- Teaching at the Right Level (TaRL). Developed by the NGO Pratham, TaRL groups children by current learning level rather than grade for part of the school day and teaches from that level. A series of randomised evaluations across several states found substantial learning gains when the approach was implemented as designed, and documented what was needed to scale it within

government systems (Banerjee et al., 2017). TaRL acts directly on M5, and its scaling experience shows M1 at work: gains depended on protecting dedicated time and grouping within the school timetable.

- Adaptive technology. Mindspark's gains (Section 4.3) were largest in relative terms for the weakest students, whose learning in the control group was close to zero (Muralidharan et al., 2019).

Both interventions change the structure of instruction, not just its content. That is the pattern the account predicts will succeed.

6. Testable Propositions

The account yields propositions that could be supported or refuted by future evidence:

- P1 (structure over content). Reforms that change the organisation of instruction (grouping, pacing, time) will produce larger learning gains than reforms that change curriculum content alone, holding cost constant.
 - P2 (absorption). Curriculum reforms introduced without corresponding changes to high-stakes examinations will show smaller effects on classroom practice than reforms accompanied by aligned assessment changes.
 - P3 (multitask distortion). Strengthening incentives on narrowly measured outcomes will raise those outcomes while reducing performance on correlated but unmeasured outcomes, detectable with low-stakes audit tests.
 - P4 (level-matched instruction). The benefit of level-matched instruction will be greatest where within-class variance in learning levels is largest.
 - P5 (credential pressure). The intensity of examination-oriented instruction and private coaching will increase with the labour-market return to credentials relative to measured skills.
 - P6 (spending effectiveness). Additional spending will raise learning more when directed at instructional time and level-matched teaching than when directed at infrastructure or administration.
-

7. Implications for Institutional Redesign

The account suggests five design principles:

1. Target structure, not only content. Protect time for level-based grouping; decouple progression from age where feasible.
2. Align assessment with the reform. Change high-stakes examinations together with curricula, and use low-stakes diagnostic assessment to guide teaching.
3. Use balanced incentives. Pair any outcome-based incentive with audit measures of unmeasured outcomes, and avoid strong incentives on narrow metrics.

4. Deploy technology for personalisation. Invest in tools that adapt instruction to each learner's level, and evaluate them rigorously; do not simply digitise cohort instruction.
 5. Reduce the signalling load on schools. Where possible, give employers and higher-education institutions direct measures of skill, reducing reliance on credentials as proxies.
-

8. Limitations

- Conceptual, not empirical. The paper synthesises existing evidence and does not test the propositions itself.
 - Selective, narrative review. Sources were chosen for relevance rather than through a systematic search; the evidence base is weighted toward the United States and India.
 - Contested mechanisms. The relative importance of signalling versus human capital, and the net effect of accountability, remain actively debated; the account does not settle them.
 - Heterogeneous systems. “Global education systems” differ widely. The mechanisms are proposed as common tendencies, and their strength will vary by country and level of education.
-

9. Conclusion

Education systems are not simply failing; they are stable, because their core practices are held in place by mutually reinforcing mechanisms: an institutional grammar that defines what school is, the signalling value of credentials, high-stakes assessment, multitask incentive problems and cohort-based instruction. Reforms that change content while leaving these mechanisms intact tend to be absorbed. The evidence also shows that the mechanisms can be weakened. Money spent well, accountability designed with its distortions in mind, and above all interventions that match instruction to each learner's level have produced large, measurable gains, including in India's government schools. The design challenge for the coming decades is to build institutions in which the structure of schooling serves learning, not only sorting.

References

ASER Centre. (2025). *Annual status of education report (rural) 2024*. Pratham Education Foundation. <https://asercentre.org/>

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of*

Economic Perspectives, 31(4), 73–102.

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

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 7(3), 375–385.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. Basic Books.

Brown, D. K. (2001). The social sources of educational credentialism: Status cultures, labor markets, and organizations. *Sociology of Education*, 74, 19–34.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90.

Caplan, B. (2018). *The case against education: Why the education system is a waste of time and money*. Princeton University Press.

Collins, R. (1979). *The credential society: An historical sociology of education and stratification*. Academic Press.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Harvard University Press.

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.

Dee, T. S., & Jacob, B. A. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446.

Gatto, J. T. (1992). *Dumbing us down: The hidden curriculum of compulsory schooling*. New Society Publishers.

Gorard, S. (2014). The link between Academies in England, pupil outcomes and local patterns of socio-economic segregation between schools. *Research Papers in Education*, 29(3), 268–284.

Guskey, T. R. (2010). Lessons of mastery learning. *Educational Leadership*, 68(2), 52–57.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, 113(485), F64–F98.

- Holmström, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24–52.
- Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics*, 131(1), 157–218.
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3), 843–877.
- Koretz, D. (2017). *The testing charade: Pretending to make schools better*. University of Chicago Press.
- Melton, J. V. H. (1988). *Absolutism and the eighteenth-century origins of compulsory schooling in Prussia and Austria*. Cambridge University Press.
- Ministry of Human Resource Development. (2020). *National Education Policy 2020*. Government of India.
- Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.
- Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4), 1426–1460.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Harvard Education Press.
- Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Center for Global Development.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. Basic Books.
- Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking*. Wiley.
- Sahlberg, P. (2015). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* Teachers College Press.
- Scafidi, B. (2012). *The school staffing surge: Decades of employment growth in America's public schools*. Friedman Foundation for Educational Choice. <https://eric.ed.gov/?id=ED536674>
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355–374.
- Tyack, D. (1974). *The one best system: A history of American urban education*. Harvard University Press.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Harvard University Press.

- Tyack, D., & Tobin, W. (1994). The “grammar” of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31(3), 453–479.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- World Bank. (2018). *World development report 2018: Learning to realize education's promise*. World Bank.