DARE-RC
DATA AND RESEARCH IN EDUCATION RESEARCH CONSORTIUM
جرأتِ تحقیق ملے

UK International Development
Partnership | Progress | Prosperity

Predictive Modelling for Evidence-Informed Enrolment Strategies: Transitioning from Retrospective Reporting to Precision Governance

Evidence from the DARE Research Consortium

Policy Brief

The Equity Equation Project











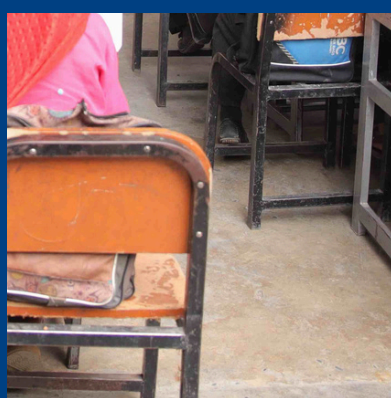

Authors
Dr Farah Nadeem,
Zariya Mushtaq

DARE-RC
DATA AND RESEARCH IN EDUCATION RESEARCH CONSORTIUM
جرأتِ تحقیق ملے

UK International Development
Partnership | Progress | Prosperity

ABOUT
This policy brief was produced by the DARE Research Consortium. It summarises the findings from a full report (forthcoming) detailing research conducted as part of the DARE-RC portfolio, and was authored by the researchers of that report. This brief forms a part of a series to provide actionable evidence to policymakers in Pakistan to improve education.

DISCLAIMER
The views expressed in this document are those of the authors and the DARE Research Consortium and do not necessarily reflect those of the FCDO or the UK Government.

PUBLICATION DATE
October, 2026

DESIGN & EDITING
Sam Wilson, Amina Batool (Wyeside Education)

Scan to find out more

DARE Research Consortium

DARE-RC is Pakistan's largest education research programme, generating rigorous, actionable evidence to improve education policy and practice.
DARE's 44 research studies and scoping papers offer critical insights on inclusion, teaching and learning, governance, data, accountability, and system resilience.

© DARE Research Consortium

AT A GLANCE DARE RESEARCH CONSORTIUM · POLICY BRIEF SUMMARY

Predictive Modelling for Evidence-Informed Enrolment Strategies: Transitioning from Retrospective Reporting to Precision Governance

Dr Farah Nadeem, Zariya Mushtaq · The Equity Equation Project

Methodology

- **Phase 1:** synthesis of PSLM/MICS national survey data, plus an 800-household primary survey and Educational Journey Mapping in Lahore Division, to train the simulation model.
- **Phase 2:** calibration using Explainable AI (XAI), co-designed with the Pakistan Institute of Education and provincial departments.

Key Figures

26.2M
children currently out of school

33%
of school-age population out of school

800
households surveyed in Lahore Division

1.5km
distance beyond which school quality stops mattering

20%
of willing households blocked by missing B-Forms

12yrs
age after which boys' re-enrolment window closes

Findings

1
Rural Infrastructure Gaps
In high-deprivation districts like Rajanpur and Kohistan, a functional toilet and boundary wall is the strongest predictor of enrolment.

2
Urban Digital Divide
In metropolitan centres such as Lahore and Islamabad, digital assets – internet access and computer literacy – serve as the new gatekeepers to education.

3
Disability and Education
Disability's negative impact is substantially mitigated where the household head has tertiary education; low-literacy rural households face a double disadvantage.

4
Shifting Gender Dynamics
In conservative rural areas, gender is the main barrier to school. In cities, poverty is a much stronger factor than gender alone.

5
Gendered Re-enrolment Windows
The window for boys to return to school closes after age 12; for girls it closes at puberty, driven by safety concerns and marriage expectations.

6
Hidden Administrative Barriers
Missing birth registration or B-Forms emerged as a significant hidden obstacle for 20% of mapped households.

Policy Implications

- **Policy design must shift into a predictive phase, testing interventions in silico before public resources are committed.**
- **For girls, infrastructure is about safety and privacy, not just logistics – this needs a policy focus on perceived safety.**
- **Distance cancels out quality: once a school is more than 1.5 km away, its quality no longer matters to parents.**
- **Effective data governance is achievable by augmenting existing survey machinery, not building expensive new systems.**

Recommendations

SHORT-TERM · 0–2 YRS
Institutionalise the model
- Mandate predictive modelling before approving development schemes.
- Standardise disability data using Washington Group questions.
- Integrate siloed data into one Decision Support System.

MEDIUM-TERM · 2–5 YRS
Target the interventions
- Launch GIS-informed safe school transit for adolescent girls.
- Deploy mobile classrooms for at-risk and dropped-out children.
- Simplify B-Form and birth registration requirements.

LONG-TERM · ONGOING
Sustain the feedback loop
- Augment existing surveys rather than build new systems.
- Institutionalise co-design sessions with parents and local leaders.

Key Messages

A central governance challenge in Pakistan's education landscape is the inability to use existing data to make targeted, evidence-based policy decisions. This can be improved using predictive models, bringing together datasets, and using innovations in data governance – helping the government shift from reactive reporting to precision governance.

Children are left out of school for a mix of reasons that vary from one neighbourhood to the next, so generic 'one-size-fits-all' policies will not work. In rural areas, girls are especially affected, and infrastructure gaps like missing toilets and boundary walls matter, whereas in urban areas, poverty and low access to digital devices are the bigger obstacles.

The Equity Equation Project (EEP) has developed a model that allows policymakers to test potential interventions by predicting how families will respond to a particular policy, before they commit public funds. The predictive model adapts existing government datasets (PSLM, MICS) and uses 'Explainable AI' so officials can explore the thinking behind each recommendation.

The model's simulation data found that distance cancels out quality: once a school is more than 1.5 km away, its quality no longer matters to parents. The model also shed light on the impact of disability on enrolment and highlighted several behavioural patterns, e.g., child labour as a coping strategy that follows dropout, and the hurdles for households lacking documentation like B-Forms.





Introduction

Pakistan currently grapples with a severe national education crisis, with approximately 26.2 million children, nearly one-third of the school-age population, remaining outside the formal schooling system. While the government has declared an Education Emergency, the National Education Policy Development Framework (2024) notes that the system's primary failure is not merely a lack of data, but the inability to use that data to create targeted policies. Currently, decision-making often relies on broad guesses that fail to capture the local, real-life reasons why a family keeps a child at home.

26.2M
CHILDREN OUT OF SCHOOL
Nearly one-third of the school-age population

> ***Enrolment is not driven by education policies alone, but shaped by poverty, gendered expectations, and safety concerns.***

The Equity Equation Project (EEP) addresses this gap by moving away from simply reporting past failures to predicting future solutions. By placing the family unit at the centre of analysis, the project's simulation model recognises that enrolment is not driven by education policies alone; rather, it is shaped by a mix of poverty, gendered expectations, and safety concerns that define the lived reality of marginalised children.

The overarching objective of this project is to strengthen evidence-based planning. By integrating government records with local household and geospatial (GIS) data, the EEP allows policymakers to simulate how different families respond to specific policies, such as building a new school or providing transport, before any funds are committed. This ensures that the limited government resources are spent where they will have the greatest impact on Pakistan's most vulnerable children.

Methodology

The study utilises a rigorous mixed-methods design that integrates quantitative, qualitative, and geospatial data into a cohesive modelling framework, conducted in two primary phases.

Phase 1: Conceptualisation and Development: The team synthesised secondary data from national surveys (PSLM, MICS) to identify macro-level predictors of enrolment. This was augmented by a primary survey of 800 households in the Lahore Division and qualitative Educational Journey Mapping to capture micro-level behavioural insights, used to train a simulation model.

Phase 2: Calibration and Engagement: The project leveraged Explainable AI (XAI) techniques[1] that unpack how different factors interact to influence school enrolment, allowing non-technical officials to see the logic behind every prediction.

Co-Design Process: Through a collaborative approach involving the Pakistan Institute of Education (PIE) and provincial departments, the model was calibrated to administrative realities to facilitate institutional adoption.

Limitations: Primary data collection was concentrated in the Lahore Division, which may limit generalisability to other provincial terrains. Successful adoption also assumes functional planning units and data interoperability capacity, which currently varies across provinces.

Key Findings

The EEP project shows that children miss out on school for many different, overlapping reasons. This means a single, 'one-size-fits-all' policy will not work for every community. By integrating administrative, survey, and geospatial data, the study uncovers intersecting determinants of enrolment varying by geographical location that traditional models often overlook.

Structural Determinants and the Geography of Deprivation

Predictive modelling reveals a distinct hierarchy of exclusion that shifts based on geographic context.

Rural Infrastructure Gaps: In high-deprivation districts like Rajanpur and Kohistan, fundamental infrastructure deficits are the primary binding constraints. The presence of a functional toilet and boundary wall is the strongest predictor of enrolment, validating the theory that inadequate facilities actively deter participation.

Urban Digital Divide: In metropolitan centres such as Lahore and Islamabad, exclusion is modernised. Digital assets, including internet access and computer literacy, serve as the new gatekeepers to educational opportunities.

Intersectional Drivers and Navigational Capital

Disability and Education: While disability is a significant universal barrier, its negative impact is substantially mitigated in households where the head has a tertiary education. Educated parents possess the navigational capital to advocate for their children, buffering them against structural exclusion. In contrast, children with disabilities in low-literacy, rural households face a double disadvantage.

Shifting Gender Dynamics: In conservative rural areas, gender is the main barrier to school. In cities, however, poverty is a much stronger factor in keeping children out of school than gender alone.

Behavioural Patterns and the Re-enrolment Window

Gendered Windows: The opportunity for boys to return to school closes after age 12 due to rising labour-market pressures. For girls, the window closes at puberty, primarily due to safety concerns and marriage expectations.

Hidden Administrative Barriers: Documentation hurdles, such as a lack of birth registration or B-Forms, emerged as a significant hidden obstacle for 20% of mapped households.

Cost vs. Labour: Data confirms that economic constraints are the primary driver of dropout. Child labour is overwhelmingly a coping strategy that follows dropout rather than causes it.

Data Silos and Integration Opportunities

The research process highlights a data-rich but information-poor environment. While Pakistan possesses vast repositories of survey and administrative data, these datasets remain siloed. However, this study demonstrates that by augmenting existing EMIS and survey datasets with targeted behavioural indicators, simulation modelling can effectively inform responsive, targeted governance.

Policy Implications

The findings of EEP necessitate a fundamental shift in Pakistan's approach to the Education Emergency, moving from reactive, ad-hoc administration towards proactive policy engineering. Monolithic national mandates, and even sub-national policies, are inherently ill-equipped to address the diverse drivers of exclusion. Policy design must transition into a predictive phase where interventions are tested in silico to anticipate household responses before public resources are committed.

Furthermore, the study reveals that distance cancels out quality, which has profound implications for infrastructure planning. This underscores that school construction alone is insufficient; infrastructure must be integrated with cross-sectoral transport and safety policies to be effective. For girls in particular, infrastructure is about safety and privacy; deficits like the absence of boundary walls or functional toilets are interpreted by parents as signals of cultural and physical insecurity, necessitating a policy focus on perceived safety rather than just objective quality.

Finally, the project demonstrates that effective data governance is achievable even under significant resource constraints. Rather than creating expensive new data systems, the government can achieve advanced predictive capabilities by augmenting existing national survey machinery with targeted behavioural and perception-based questions. This provides a sustainable roadmap for systemic accountability, ensuring that resource allocation is driven by objective need rather than political expediency.

> ***Distance cancels out quality: once a school is more than 1.5 km away, its quality no longer matters to parents.***

> ***Resource allocation should be driven by objective need rather than political expediency.***

Recommendations

To translate these insights into systemic change, EEP offers actionable recommendations grounded in empirical findings and aligned with provincial and federal priority areas.

Institutionalising Predictive Tools and Inclusive Data · Provincial Governments · Short-term, 0–2 years

- **Institutionalise Predictive Modelling:** Provincial Planning departments should mandate the use of computational modelling to simulate the incremental impact of interventions before approving development schemes.
- **Standardise Disability Data Across EMIS Units:** All EMIS units should adopt the Washington Group Short Set on Functioning to ensure children with disabilities are properly counted.
- **Integrate Fragmented Data Systems:** Establish formal interoperability protocols to unify siloed administrative, survey, and GIS datasets into a centralised Decision Support System.

Targeted Interventions for Access and Retention · Provincial Governments · Medium-term, 2–5 years

- **Launch GIS-Informed Safe School Transit Programmes:** Target transport initiatives to make schools accessible, addressing safety concerns for adolescent girls.
- **Launch Mobile Classrooms for At-Risk Children:** Use mobile 'fast-track' learning programmes to reach children at high risk of being 'too old' for their grade if they do not re-enrol soon.
- **Use Partnerships to Lower Schooling Costs:** Work through partnerships like Punjab Education Foundation to provide targeted material assistance (uniforms, books, transport).
- **Simplify School Entry Paperwork:** Make birth registration and B-Form requirements flexible, addressing the finding that documentation barriers block 20% of willing families.

Sustained Capacity and Feedback Loops · Federal and Provincial Governments · Long-term, ongoing

- **Adopt Frugal Innovation[2] in Data Governance:** Regularly augment existing survey machinery (PSLM/MICS/HIES) with targeted behavioural and perception-based questions.
- **Institutionalise Design-Based Feedback Loops:** Future policy iterations must involve co-design sessions with parents and local leaders.

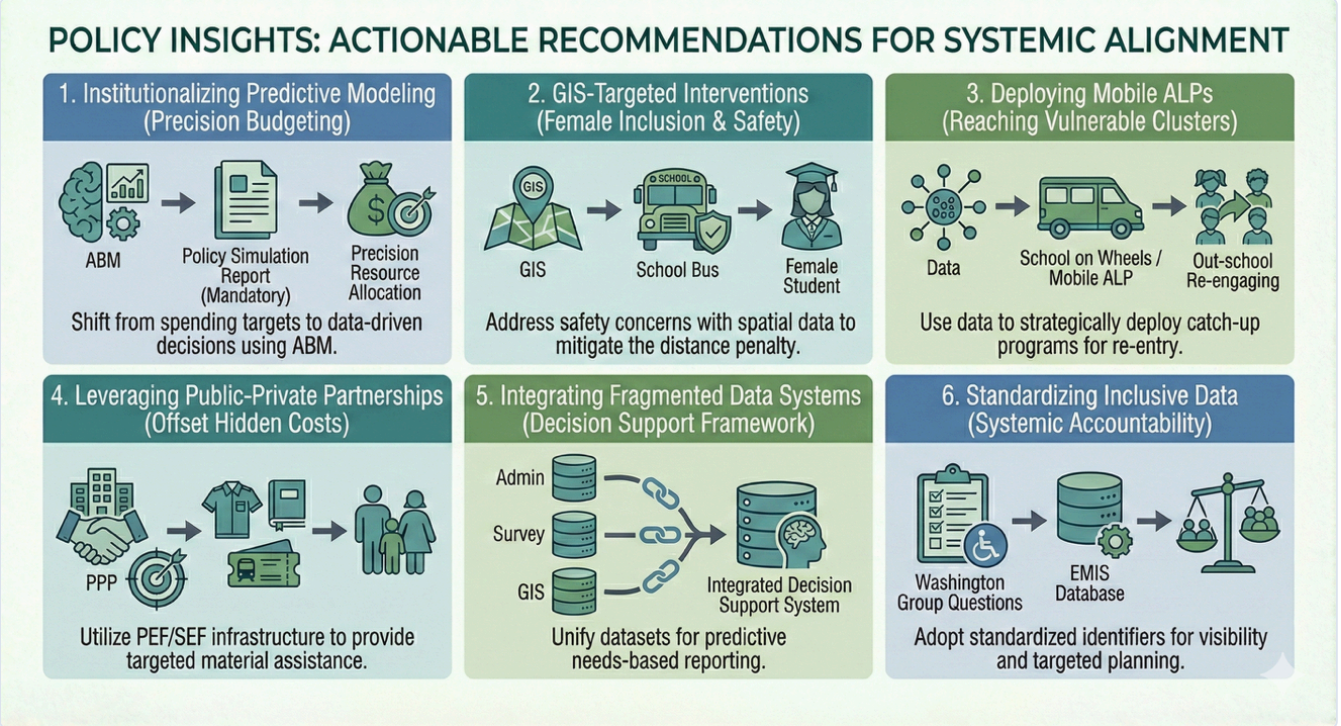

Figure 1

Policy insights – actionable recommendations for systemic alignment

[1] Explainable AI acts like an open book. Instead of just giving a final prediction, the system clearly shows why it reached a certain conclusion.

[2] A strategy that focuses on getting high-quality results using existing tools and low-cost methods, making small updates to the surveys the government already conducts rather than spending on new, expensive data systems.

Conclusion

The research undertaken as part of EEP demonstrates that the persistent crisis of educational exclusion in Pakistan is not merely a resource problem; rather, it is a design problem rooted in fragmentation and the under-utilisation of evidence. The findings confirm that the reasons children stay out of school change from one neighbourhood to the next. By recognising these local differences, the government can finally move away from generic policies and start using precision planning to ensure every child is reached.

Engaging with these findings is critical to solving the data-rich but information-poor environment that currently hinders responsive governance. By institutionalising predictive tools alongside frugal innovation in data governance, the government can transition from reactive reporting[3] to precision governance.[4] This paradigm shift ensures that a child's objective needs, rather than political expediency, drive the future of education reform in Pakistan.

[3] This is the current 'data-rich but information-poor' status quo. The government collects massive amounts of data, but it is usually used to write reports about what went wrong in the past.

[4] Instead of applying one broad policy to an entire province, specific, micro-level barriers are identified in each locality and policies are tailored accordingly.

For More Information

This research was carried out under the DARE-RC project. The DARE-RC project is funded by UK International Development from the UK government. Its implementation is led by Oxford Policy Management in partnership with its consortium partners.

Authors: Dr Farah Nadeem, Ms Zariya Mushtaq, LUMS
Report Date: January 2026
This policy brief is based on findings from the report titled: *Equity Equation Project*
More information about the DARE-RC Programme: https://darerc.org/

DARE-RC
DATA AND RESEARCH IN EDUCATION RESEARCH CONSORTIUM
جرأتِ تحقیق ملے

UK International Development
Partnership | Progress | Prosperity

Scan to find out more

DARE Research Consortium

DARE-RC is Pakistan's largest education research programme, generating rigorous, actionable evidence to improve education policy and practice. DARE's 44 research studies and scoping papers offer critical insights on inclusion, teaching and learning, governance, data, accountability, and system resilience.

Oxford Policy Management
THE AGA KHAN UNIVERSITY
Institute for Educational Development
Sightsavers

DARE-RC is funded with UK International Development from the UK Government and its implementation is led by Oxford Policy Management in partnership with its consortium partners.