POLICY BRIEF

# Predictive Modelling for Evidence-Informed Enrolment Strategies - Transitioning from Retrospective Reporting to Precision Governance in Pakistan's Education Emergency

## The Equity Equation Project

### Evidence from the DARE-RC portfolio

By Dr Farah Nadeem, Ms. Zariya Mushtaq

## Key Messages

- **A central governance challenge in Pakistan is the inability to use existing data to make targeted, evidence-based policy decisions.** Pakistan has substantial existing data sets that can be used in decision making. This can be improved using predictive models, bringing together data sets, and using innovations in data governance. This will help the government shift from reactive reporting to precision governance of Pakistan's education emergency.

- **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.** For this, the government does not need new data. The predictive model adapts existing government datasets (PSLM, MICS), and uses 'Explainable AI' so officials can explore the thinking behind each recommendation.

- **Children are left out of school for a mix of reasons that vary from one neighborhood 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. Given these nuances, interventions must be tailored to local needs and can first be tested out in the predictive model.

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

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.

## 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.

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 center of analysis, the project's simulation model recognizes that enrollment 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 marginalized 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 utilizes a rigorous mixed-methods design that integrates quantitative, qualitative, and geospatial data into a cohesive modeling framework. The research was conducted in two primary phases to ensure technical accuracy and policy relevance:

- **Phase 1: Conceptualization and Development:** The team synthesized secondary data from national surveys: Pakistan Social and Living Standards Measurement Survey (PSLM) and Multiple Indicator Cluster Survey (MICS) to identify macro-level predictors of enrollment. This was augmented by a primary survey of 800 households in the Lahore division and qualitative Educational Journey Mapping to capture micro-level behavioral insights, which were then used to train a simulation model predicting probability of enrollment for a family with given demographic, access, and school-level factors.
- **Phase 2: Calibration and Engagement:** The project leveraged Explainable AI (XAI) techniques[1] that unpack how different factors interact to influence school enrollment, allowing non-technical officials to see the logic behind every prediction.

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.

- **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:**

- **Geographic Scope:** While this conceptualization study provides high-fidelity insights, primary data collection was concentrated in the Lahore Division, which may limit immediate generalizability to vastly different provincial terrains.
- **Implementation Assumptions:** Successful adoption assumes existing functional planning units and data interoperability capacity, which currently varies across provinces; uptake planning in regions like KP and Balochistan must account for these varying baselines.

Crucially, the methodology demonstrated that existing national data pipelines can be enhanced with targeted behavioral questions to support robust simulations without requiring extensive additional investment.

## 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 enrollment varying by geographical location that traditional models often overlook.

### *Structural Determinants and the Geography of Deprivation*

Predictive modeling 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. Specifically, the presence of a functional toilet and boundary wall are the strongest predictor of enrollment, validating the theory that inadequate facilities actively deter participation.

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

### *Intersectional Drivers and Navigational Capital*

Intersectional analysis highlights how compounding disadvantages create deep exclusion:

**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*

Qualitative evidence uncovers distinct, time-bound behavioral windows for re-entry:

**Gendered Windows:** The opportunity for boys to return to school closes after age 12 due to rising labor-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. Labor:** Data confirms that economic constraints are the primary driver of dropout. Child labor is overwhelmingly a coping strategy that follows dropout rather than causing 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 behavioral indicators, simulation modeling 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. The primary implication of this research is that 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**. This approach is particularly critical in a constrained fiscal environment, as it allows for precise targeting of high-impact interventions, such as sanitation in rural Balochistan versus digital literacy in urban Punjab.

**Furthermore, the study reveals that distance cancels out quality, which has profound implications for infrastructure planning.** Simulation data demonstrates that once a school is more than 1.5km away, its quality no longer matters to parents. 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 not merely logistical gaps but 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, such as PSLM and MICS, with targeted behavioral and perception-based questions**. By adopting Explainable AI, policymakers can scrutinize algorithmic recommendations and verify the extent to which they reflect real conditions. This will help build the institutional confidence needed to act on them. This provides a sustainable roadmap for systemic accountability, ensuring that resource allocation is 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.

1. **Institutionalizing Predictive Tools and Inclusive Data by Provincial Governments (Short-term 0-2 years)**
    - **Institutionalize Predictive Modeling:** Provincial Planning departments should mandate the use of computational modeling to simulate the incremental impact of interventions before approving development schemes. This allows for policy engineering through *in silico* testing of interventions.
    - **Standardize Disability Data Across EMIS Units:** To fix the visibility gap identified in the findings, all EMIS units should adopt the Washington Group Short Set on Functioning to ensure children with disabilities are properly counted. Without this data, planning systems cannot identify/prioritize children who face multiple disadvantages, such as disability alongside low household literacy
    - **Integrate Fragmented Data Systems:** Establish formal interoperability protocols to unify siloed administrative, survey, and GIS datasets into a centralized Decision Support System. This enables the government to move from retrospective reporting toward the predictive, needs-based reporting required for responsive governance.
2. **Targeted Interventions for Access and Retention by Provincial Governments (Medium-term 2-5 years)**
    - **Launch GIS-Informed Safe School Transit Programs:** Using geospatial data, the state should launch targeted transport initiatives to make schools accessible, as school quality starts to matter less to parents when schools are farther away. This

directly addresses safety concerns and cultural resistance identified as primary barriers for adolescent girls.

- **Launch Mobile Classrooms for At-Risk Children:** Use mobile "fast-track" learning programs to reach children in remote or poor areas who have already dropped out or never started school. These mobile units should use the project's map data to find groups of children who are at high risk of being "too old" for their grade if they don't re-enroll soon. This provides a flexible catch-up path for both boys and girls, ensuring they don't permanently lose their chance to enter the formal school system.
- **Use Partnerships to Lower Schooling Costs:** The government should work through partnerships such as Punjab Education Foundation and Sindh Education Foundation to help the families with hidden costs by providing targeted material assistance (uniforms, books, transport). This addresses the finding that economic hardship, not child labor, is the leading driver of children dropout.
- **Simplify School Entry Paperwork:** Launch a targeted drive to remove hidden paperwork barriers by making the birth registration and B-Form requirements for school enrollment flexible. This addresses the study's finding that documentation barriers block 20% of willing families from the system because they cannot navigate the complex or costly process of obtaining the required documents. This will allow the willing families extra time to get their children's paperwork in order while affording them the opportunity to get their children enrolled immediately.

3. **Sustained Capacity and Feedback Loops by Federal & Provincial Governments (Long-term; Continuous/Ongoing)**

    - **Adopt Frugal Innovation[2] in Data Governance:** Regularly augment existing survey machinery (PSLM/MICS/HIES) with targeted behavioral and perception-based questions. This avoids expensive new infrastructure while maintaining the high-fidelity data required for predictive modeling.

    - **Institutionalize Design-Based Feedback Loops:** Future policy iterations must involve co-design sessions with parents and local leaders to ensure simulation outputs remain grounded in lived realities and private preferences.

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

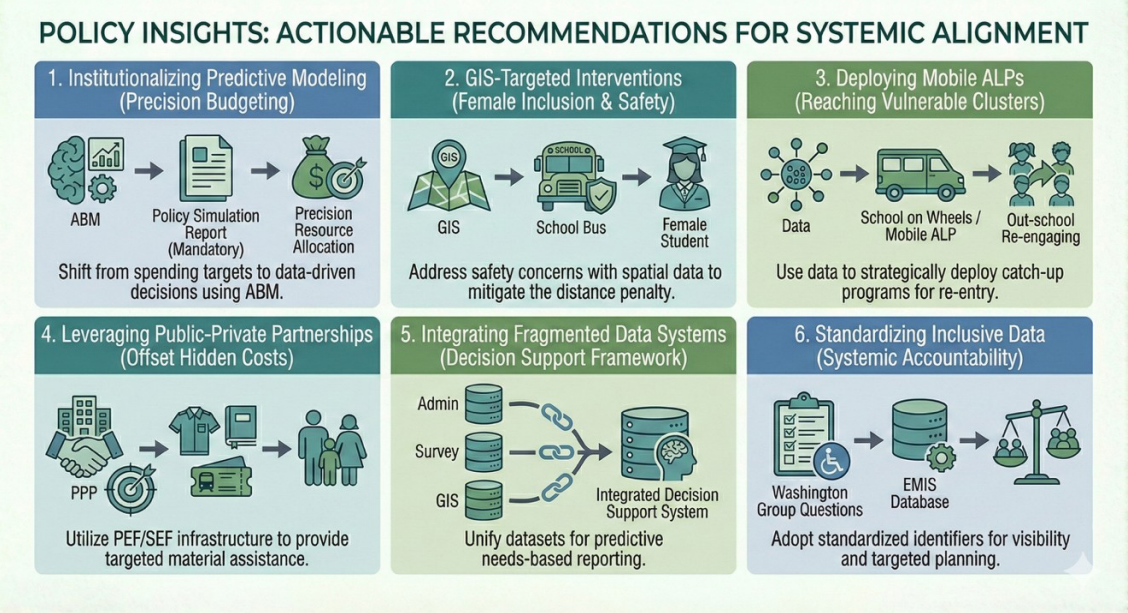


*Figure 1: Policy insights*

## 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 underutilization of evidence. Over the course of this study, we successfully moved beyond traditional descriptive analysis to pilot a predictive computational model framework that integrated administrative, demographic, and geospatial data into a coherent decision-support system. Our findings confirm that the reasons children stay out of school change from one neighborhood to the next. While a rural family might be held back by a lack of school toilets or safety concerns, an urban family might be blocked by a lack of internet access or economic constraints. By recognizing these local differences, the government can finally move away from generic policies and start using precision planning to ensure every child is reached.

Engaging these findings is critical to solve the data-rich but information-poor environment that currently hinders responsive governance. By institutionalizing predictive tools alongside frugal innovation in data governance, the government can transition from reactive reporting[3] to precision governance.[4] Further, utilizing Explainable AI (XAI) to provide interpretable, transparent insights into enrollment drivers builds the administrative trust necessary to ensure that proactive, evidence-informed planning becomes the standard for addressing Pakistan's education emergency. 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 with 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.