Female Human Capital and Children's Educational Outcomes in Indian Districts

Arunima Mandwariya & Parth Mahajan

Data Science for Social Science Research

Professor Anustubh Agnihotri

May 8, 2025

1. Introduction and Motivation

Children's performance on foundational learning in India, defined through basic literacy and numeracy, remains worryingly low. Nationally representative surveys like ASER have consistently revealed that children struggle to read simple text or perform basic arithmetic even after several years of schooling. This persistent learning crisis has generated a body of research investigating school infrastructure, teacher quality, and economic deprivation. Yet, a critically underexplored determinant lies within the household: the role of female human capital, particularly maternal education and female labor force participation (FLFP).

Maternal education is widely recognized as a key factor influencing children's educational outcomes. Educated mothers are more likely to engage in their children's education, assisting with homework, enrolling them in better schools, and advocating for their rights. Previous studies (Sunder, 2019; Afridi et al., 2016) highlight this effect, showing that maternal education enhances both the quantity and quality of a child's educational experience. In contrast, the literature on paternal education shows limited or inconsistent associations with children's academic performance.

Female labor force participation (FLFP) may also play an important role in shaping children's outcomes. Working mothers can gain greater decision-making power within the household, which may increase investment in their children's education. However, there are concerns about the trade-offs between work and time spent supporting their children's learning (Vikram et al., 2018). The interaction between FLFP and educational outcomes is complex, as it is influenced by the type of employment, socio-economic status, and access to resources.

While these mechanisms have been discussed in the literature separately, few studies have evaluated their combined effect, particularly at the district level. Most research tends to focus on household-level surveys or ethnographic studies. This research aims to fill this gap by using nationally representative data from ASER 2011, merged with Census 2011 indicators and night light data from the SHRUG dataset.

Our research question is:

In what ways does female human capital shape children's educational outcomes in Indian districts?

We answer this by constructing composite indicators for children's learning, female human capital, and contextual district-level characteristics, then estimating their relationships using both descriptive and multivariate analyses, including fixed-effects regressions.

2. Literature Review

Foundational learning remains a critical challenge in India, with only a minority of primary-grade students achieving basic reading and arithmetic proficiency (ASER, 2018). A growing body of evidence highlights parental human capital—particularly that of mothers—as a central determinant of children's learning outcomes. Numerous global and India-specific studies demonstrate that maternal education significantly enhances child academic performance. According to ASER reports, while educational attainment among rural mothers has improved in recent years, nearly half of mothers of school-aged children had never attended school as of 2016. More educated mothers are generally more engaged in their children's learning: they supervise homework, maintain higher academic expectations, and invest more in learning materials and support (Banerjee et al., 2016).

Sunder (2019), using a district-level regression discontinuity based on India's District Primary Education Programme (DPEP), finds that increases in mothers' education led to measurable improvements in their children's reading and math scores (0.19–0.18 SD), whereas gains in fathers' schooling had no comparable effect. This reinforces the hypothesis that maternal education drives intergenerational human capital formation through mechanisms beyond income—particularly through greater involvement in and value placed on children's education. Luke and Munshi (2011) similarly observed in South Indian tea plantations that higher female earnings—used as a proxy for women's intra-household bargaining power—positively affected children's educational attainment.

Female labor force participation (FLFP) adds a complementary dimension to this framework. With India's female workforce participation rate among the lowest globally (circa 21% in 2019), shifts in women's employment status have outsized developmental potential. Empirical work using India's Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) offers causal evidence linking maternal employment with improved educational outcomes. Afridi et al. (2016), using Young Lives panel data from Andhra Pradesh, show that greater female participation in MGNREGA correlates with significant increases in children's school attendance and test scores. The authors argue that these gains stem not merely from additional household income, but from enhanced maternal bargaining power and autonomy in household decision-making.

Sahoo and Kumari (2016) further demonstrate that when women work under MGNREGS, their children, especially daughters spend more time in school and attain higher grades. They identify both resource effects (increased ability to finance school-related expenses) and empowerment effects (greater maternal influence over household priorities). These findings resonate with broader international literature. For instance, Qian (2008) finds in China that policies raising women's relative income improve educational outcomes, especially for girls, while increases in male income show the opposite effect. Collectively, these studies suggest that female employment can reshape intra-household dynamics in favor of children's education.

The mechanisms through which maternal education and employment influence learning outcomes are well-documented. First, better-educated mothers are more likely to engage in their children's learning activities, promote education-oriented values, and structure home environments that support cognitive development. Second, both education and employment increase women's control over household resources and decisions. For instance, Sahoo et al. (2016) find that women who earn more through NREGA direct a greater share of household spending toward education and report higher autonomy in financial choices. Third, although maternal employment could theoretically reduce time available for childcare, the empirical evidence suggests the net effect is positive. Mothers working under NREGS, for example, often facilitate school attendance more consistently, particularly for girls, either by reallocating household labor or increasing school-related spending.

In sum, existing literature converges on the view that maternal human capital—operationalized through education and labor force participation—significantly influences child learning outcomes. While much of the existing work relies on regional or program-specific studies, national-scale, district-level analyses remain scarce. Our study contributes to this gap by analyzing cross-district variation using ASER 2011 data (on foundational learning) and SHRUG (on demographic and economic variables) to assess whether districts with higher average maternal education and female labor force participation also exhibit stronger child educational performance. By controlling for economic proxies, household assets, and district-level infrastructure, we aim to isolate the specific contribution of female human capital to foundational learning in India.

3. Data

Data Sources and Variable Construction

We use two major datasets: ASER 2011 and SHRUG to explore the relationship between female human capital and foundational learning outcomes of children aged 5–16 across Indian districts. Both datasets are from 2011, allowing for consistent temporal alignment.

Annual Status of Education Report (ASER), 2011 – Household and School-Level Datasets
ASER is India's largest citizen-led household survey that provides nationally representative data on children's learning outcomes in reading and arithmetic, collected through one-on-one assessments. In addition to education, it captures basic household characteristics such as parental education, access to infrastructure, and school enrollment. For this study, we secured access to the full household-level ASER dataset through an application to ASER, since the publicly available dataset is limited to district-level aggregates. The ASER household dataset comprises 6,04,144 individual child-level entries from rural India, providing rich information on child learning levels, maternal education, and household infrastructure. The school-level dataset includes data from 15,245 schools, covering infrastructure indicators such as availability of toilets, computers, and mid-day meals.

Socioeconomic High-resolution Rural-Urban Geographic (SHRUG) Database
The SHRUG is a high-resolution spatial dataset combining various government datasets to allow for district- and sub-district-level socioeconomic analysis. It includes restructured information

from the 2011 Population Census, economic indicators like night-time light emissions (as a proxy for urbanization), and demographic breakdowns. In our study, we use SHRUG to extract female labor force participation (FLFP) and caste composition. It also provides access to VIIRS night-light data, used as a continuous proxy for urban development and economic activity.

We restrict our analysis to 2011 due to data availability issues in later years. ASER datasets for some years were either incomplete or not publicly accessible, making longitudinal analysis infeasible. As such, we conduct a cross-sectional analysis at the district level which is the most granular common unit across all datasets.

Dependent Variable – Foundational Learning Outcomes

Our key dependent variable, children's arithmetic and reading skills, are derived from the ASER 2011 household dataset. These are assessed using standardized tests administered to children aged 5–16 at their homes. ASER's methodology allows for the classification of learning outcomes along ordinal scales.

Arithmetic Skills

Arithmetic ability is captured using a four-category ordinal variable based on the child's ability to perform specific tasks:

1. No recognition of written numbers
2. Can recognize and read numbers from 1-9
3. Can recognize and read numbers from 10-99
4. Can subtract a two-digit number from another two-digit number
5. Can divide a three-digit number by a one-digit number

Each level is coded as missing or 1.

Reading Skills

Reading ability is assessed using a five-category ordinal variable:

1. Cannot read at all
2. Can read individual letters

3. Can read common words
4. Can read a short paragraph (around three sentences)
5. Can read a one-page story

Here also, each level is coded as missing or 1.

We created ordinal indicators for children's arithmetic and reading abilities based on the ASER dataset and normalized them to a 0-1 scale. For arithmetic, a score (math_score) is assigned based on the highest level of math the child can perform, with values ranging from 0 (no skills) to 4 (highest level of skill, such as division). A similar approach is applied to the reading score (reading_score), where values range from 0 (cannot read letters) to 4 (can read a Class 2 level story). Both scores are then scaled by dividing them by the maximum possible score (4), resulting in scaled scores (math_score_scaled and reading_score_scaled) that range from 0 to 1.

Finally, a **composite score** is calculated by averaging the scaled math and reading scores, providing an overall measure of foundational learning. Missing values are handled appropriately, and any child with missing data in either reading or math has a missing composite score. This process created standardized indicators for educational outcomes, making it easier to compare and analyze the performance across different groups.

Later, using ASER's household-level weighting multipliers, we then aggregate this measure to compute the district-level average (avg_child_edu), our main outcome variable.

Key Independent Variables – Female Human Capital

We operationalize female human capital through two district-level indicators:

Maternal Education

This is measured as the average number of years of education completed by mothers of school-going children, based on the ASER household data. Given maternal education's established influence on child health and learning, this serves as a proxy for intergenerational transmission of human capital. We also categorize maternal education into five groups: very low, low, medium, high, and very high, based on cutoffs provided in the ASER survey manual.

Female Labor Force Participation (FLFP)

FLFP is calculated as the share of working-age women (15–64 years) engaged in economic activity (main or marginal work), using data from the Population Census Abstract 2011 (via SHRUG). As direct district-level FLFP measures were unavailable, we constructed this variable by dividing the number of working females by the total female population per district. This indicator captures female economic agency and public participation.

Control Variables and Indices

To account for confounding influences, we control for a broad range of household, school, and district-level variables:

Household Characteristics Index

A composite score representing household material well-being, constructed from ASER household-level data. It includes:

- TV ownership (hh_tv)
- Mobile phone (hh_mobile)
- Electricity connection (hh_electricity_conn)
- Toilet access (hh_toilet)
- Housing type (hh_type: kaccha or pucca)

After district-level consolidation, this dataset included 560 district-level observations.

School Infrastructure Index

Constructed using ASER school-level data, this index captures availability of:

- Mid-day meal (midday_meal_in_school)
- Boundary wall (boundary_wall)
- Computers and functionality (computer_in_school, computer_in_school_usable)
- Operational toilets for girls and boys (girls_toilet_in_school, boys_toilet_in_school)

After cleaning and aggregating, this dataset yielded 559 district-level observations.

Urbanization Proxy: We use VIIRS annual average night-time light intensity (from SHRUG) to classify districts as urban (mean > 0.95) or rural. This provides a consistent proxy for local economic activity and infrastructure access.

Demographics – Caste: SC and ST shares are taken from the 2011 Census Abstracts, as proportions of the total district population.

Paternal Education: While some literature finds paternal education has limited impact, we included average paternal education (from ASER household data) as a control to isolate maternal effects more precisely.

All datasets were consolidated at the district level using the variable district_name. The final merged dataset comprises 642 district-level observations, providing extensive geographic coverage across India.

All datasets were consolidated to district level and merged at the same level by variable "district_name". The final dataset comprises **642 observations - each signifying a district**, ensuring broad geographic coverage and enabling us to investigate **inter-district variation** in both female capital and child learning outcomes.

Data Cleaning and Challenges

To construct our final dataset for analysis, we worked with three primary datasets: the 2011 ASER household-level dataset, the 2011 ASER school-level dataset, and the 2011 Population Census Abstract at the district level. Since these datasets were collected independently, considerable effort was required to harmonize them for analysis. District names were standardized across datasets by correcting case format, trimming extra spaces, and resolving minor inconsistencies to enable successful merges.

Missing values, particularly in key variables such as maternal education and household amenities, led to the exclusion of certain districts from our analysis. To maintain the robustness and internal validity of our estimates, we opted for a listwise deletion approach, in which districts with any missing values in the variables of interest were excluded from the final sample. As a result, the number of usable observations was reduced from 642 to 435 districts.

Several variables from the ASER household dataset, particularly those capturing household-level assets (e.g., ownership of a television), were re-coded into binary indicators to enable consistent interpretation across models (e.g., 1 = owns a TV; 0 = does not). Additionally, the Population Census Abstract data required the use of external key files to merge state and district identifiers accurately, as naming conventions occasionally differed between sources.

All data cleaning, transformation, and merging steps were conducted in R using a reproducible R Markdown (.RMD) file, with clearly documented code chunks for each stage of processing. The final cleaned and merged dataset, named merged2_clean, contains 435 district-level observations and serves as the basis for all subsequent analyses.

Application of Course Methods:

Our project builds extensively on skills and concepts introduced throughout the course.

- **Data Aggregation using group_by(), mutate(), and summarise()**
  We used these functions to compute district-level summaries from individual and household-level data. For example, we calculated average maternal education and female labor force participation rates by grouping and aggregating relevant variables, enabling a smooth transition from microdata to district-level analysis.
- **Index Construction using rowMeans()**
  To capture multi-dimensional concepts like household wealth or access to amenities, we constructed composite indices by averaging binary indicators (e.g., ownership of a TV, fridge, or mobile phone). This technique allowed us to reduce complexity while retaining key patterns in the data.
- **Data Visualization: Boxplots, Heatmaps, and Spatial Maps**
  We used a variety of visual tools to explore patterns in the data:
  - Boxplots to examine the distribution of mother's education and child education outcome.
  - Heatmap/Correlation matrix to identify correlation patterns between variables.
  - Spatial maps to visualize regional disparities and geographic clustering of key indicators, such as maternal education and female labour force participation.

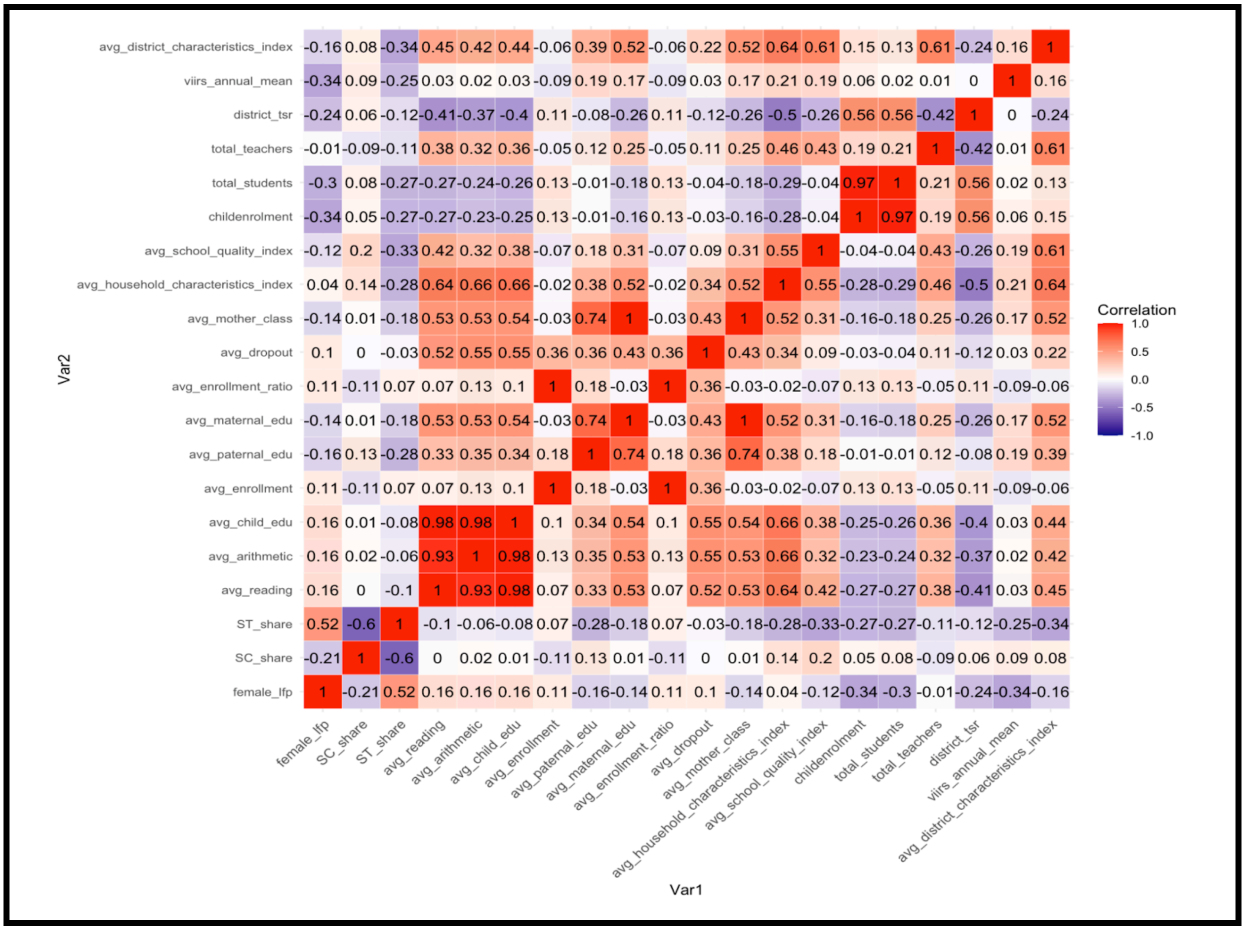
- **Regression Modeling using lm()**
  To quantify relationships between female human capital and educational outcomes, we used the **lm()** function to estimate baseline linear models. This helped establish preliminary associations between our variables of interest.
- **Fixed Effects Estimation using felm()**
  To control for unobserved heterogeneity at the state level (e.g., differences in policy environments or cultural norms), we used the **felm()** function from the *lfe* package. This allowed us to include fixed effects in our regression models and obtain more robust estimates.

4. Findings

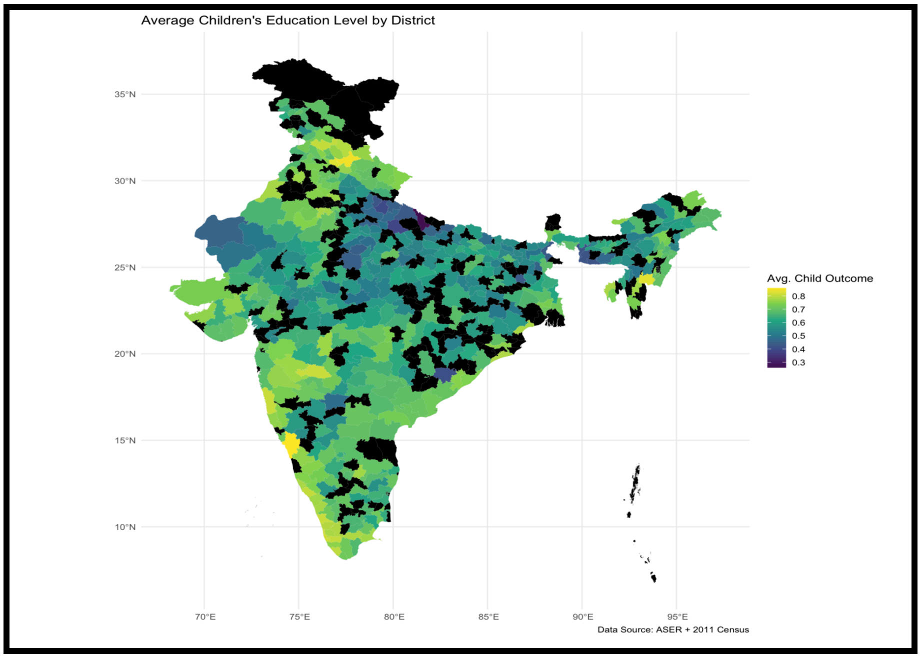
4.1 Descriptive Statistics & Correlations

Variable	Mean	SD	Min	Max
Child FLN Index	45.2%	12.1	18.3%	78.9%
Female Literacy Rate	57.5%	15.4	30.2%	89.7%
Female LFPR	25.8%	8.9	10.4%	44.5%
Maternal Education (years)	4.1	1.8	1.2	8.5
Household Characteristics Index	0 (std.)	1.0	-2.1	3.4
School Quality Index	0 (std.)	1.0	-2.8	2.9
District Infrastructure Index	0 (std.)	1.0	-3.1	4.1

Correlation matrix between all variables:

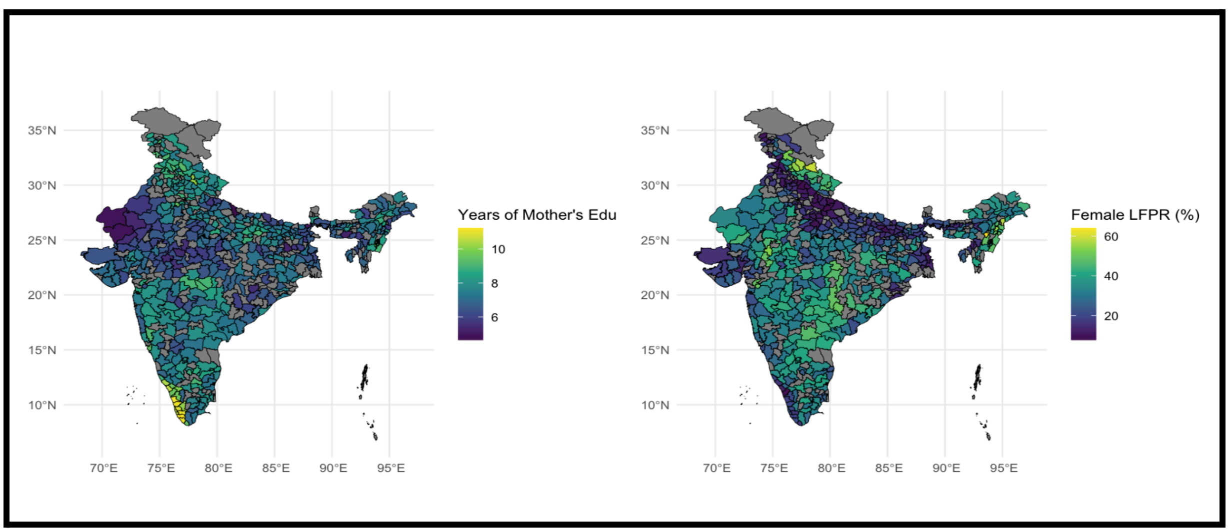


This correlation matrix reveals key relationships between educational and demographic factors. Strong positive associations exist between academic achievements (child FLN: $r \approx 0.60$, Household index & child FLN: $r \approx 0.56$, Maternal education & child FLN: $r \approx 0.48$), parental education levels, and the positive influence of household characteristics on school quality and student performance. Conversely, higher dropout rates correlate negatively with educational success (Total enrolment rates & FLN: $r \approx -0.25$). This may be suggesting quantity of enrolment not aligned with foundational learning. Notably, areas with larger Scheduled Tribe and Scheduled Caste populations may exhibit lower average educational outcomes. Many other factors show weak linear relationships, indicating independence or more complex interactions. This overview provides a concise snapshot of potential connections within the data, warranting further investigation into underlying causes.



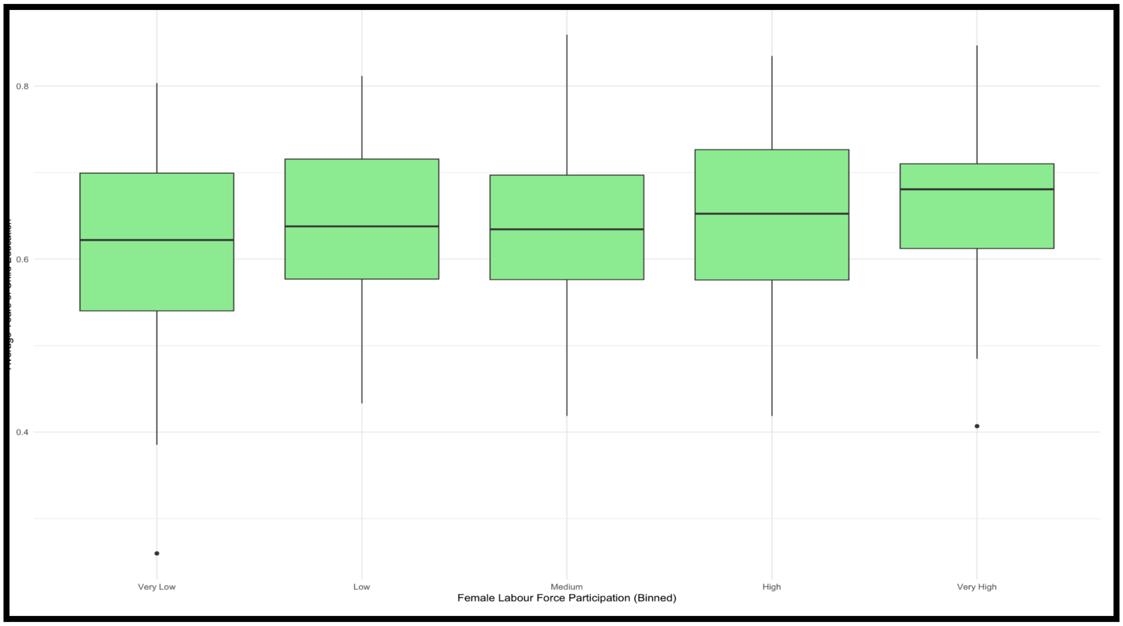
A distribution across regions can be seen with southern districts performing significantly better, corroborating with the general perception about the status of education in the South and particularly dark patch appearing over Bimaru states.

Region-wise variations of years of mother's education and Female Labour Force Participation.



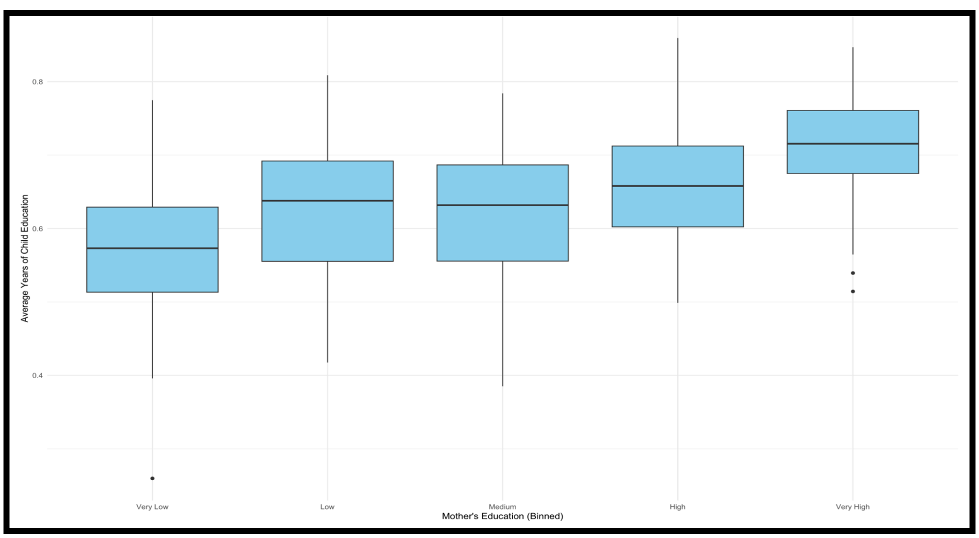
1. While Rajasthan does poorly on maternal education, it does surprisingly well in FLFPR. This can likely be explained by the type of labour - skilled/semi-skilled.)
2. Northern Plains show moderate years of mother's education, the female labor force participation in these areas is starkly low.
3. Another significant thing is the contrast of very high levels of mother's education and very low level of female labour force participation in Kerala.

Boxplots of Child FLN Scores across different levels of LFPR in a district

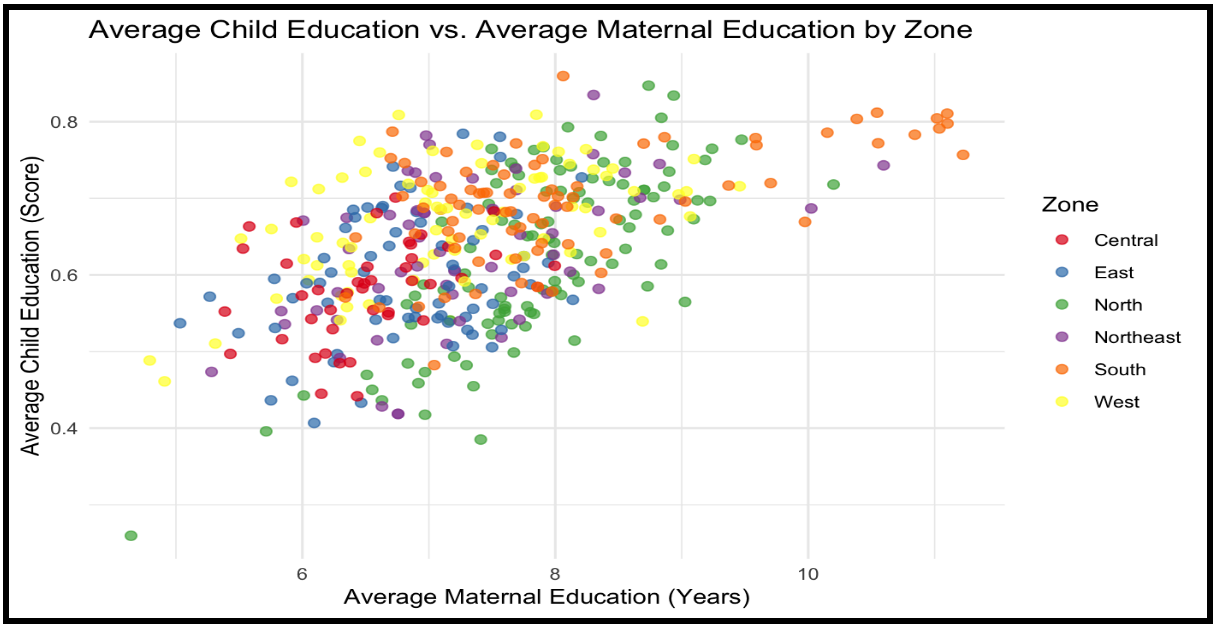


Median Education scores don’t seem to vary greatly across different levels of FLFP, which is contrary to the literature we’ve studied, prompting us to pointedly highlight it as an IV.

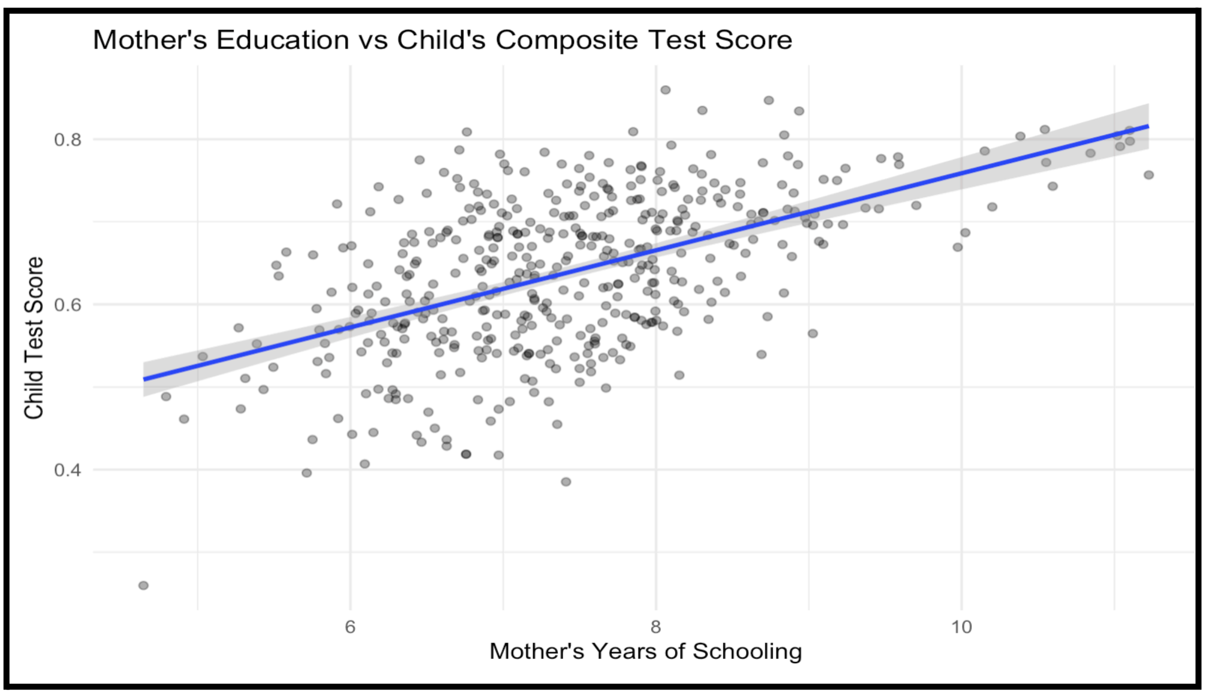
Boxplots of Child FLN Scores across different levels of Maternal Edu in a district



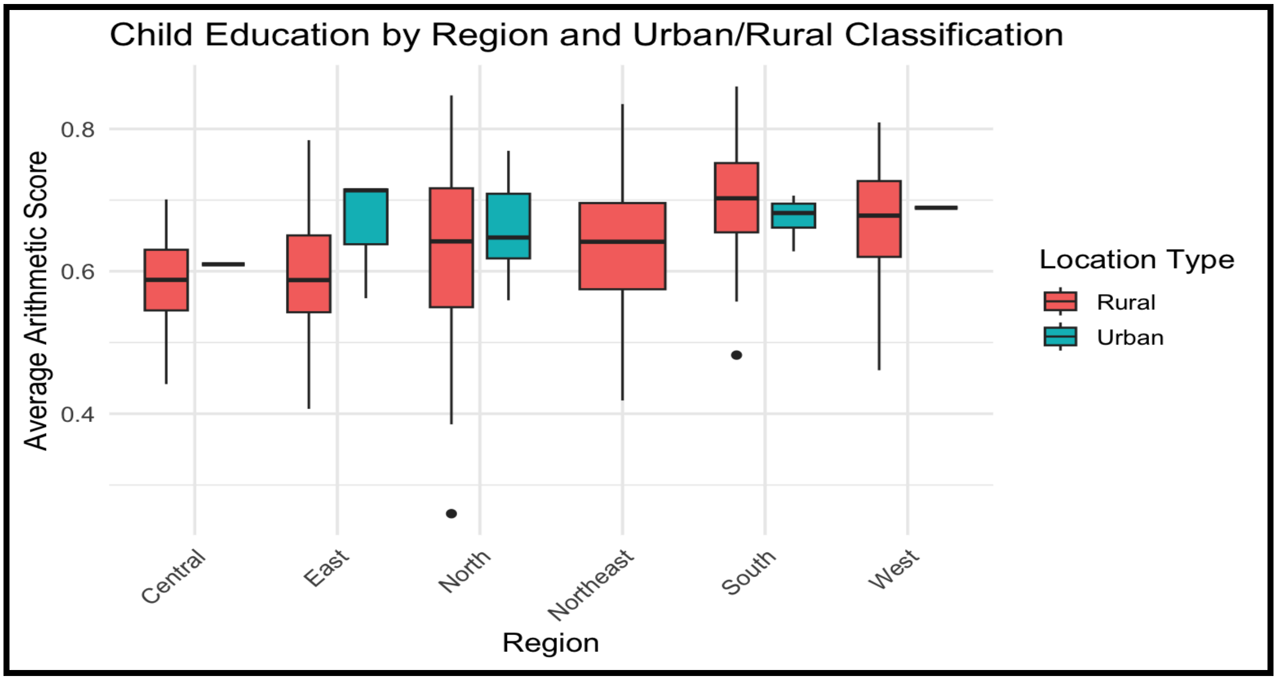
There is a non linear relationship between maternal education and child FLN scores which intuitively makes sense in that children of mothers with PG and higher education will be disproportionately better positioned and gives us confidence about the relevance of our IV



This scatterplot plots the average child education and average maternal education consolidated region-wise (refer Appendix 1). Each dot represents a district. It can be observed that the Southern states are on the extreme high ends of both the parameters.



This scatter plot shows a positive correlation between a mother's years of schooling and her child's composite test score. As the mother's education increases from about 5 to 12 years, the child's test score tends to rise from around 0.4 to 0.8.



We typically see lower variation on Urban districts, although their FLN score medians are not consistently better than their Rural counterparts. This inconsistency makes for an anomalous discovery in our final regression coefficient for Urbanity

Our descriptive analysis revealed the following:

- Districts with high avg_child_edu were clustered in southern states like Kerala, Tamil Nadu, and Karnataka. BIMARU states (Bihar, MP, UP, Rajasthan) lagged significantly.
- A strong positive trend was observed between maternal education and avg_child_edu. Boxplots showed that children of mothers in the highest education quintile had scores ~0.6–0.7, compared to ~0.3 for those in the lowest quintile.
- Female literacy and FLFP were also positively correlated with child learning. However, night-time lights—a proxy for urbanization—had a weak or even negative relationship.
- This high degree of baseline variation across states and regions - particularly North v/s South and BIMARU v/s the rest, prompts us to check for district and state fixed effects in our regression.

Regression Results: We present results from four OLS models with Incremental Adjustments:

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.2882 ***	0.2084 ***	0.2751 ***	0.2735 ***
	(0.0266)	(0.0286)	(0.0358)	(0.0359)
avg_maternal_edu	0.0471 ***	0.0502 ***	0.0307 ***	0.0315 ***
	(0.0035)	(0.0034)	(0.0048)	(0.0049)
female_lfp		0.0020 ***	0.0013 ***	0.0012 ***
		(0.0003)	(0.0003)	(0.0004)
viirs_annual_mean			-0.0050 *	-0.0052 *
			(0.0025)	(0.0026)
avg_paternal_edu			-0.0069	-0.0068
			(0.0051)	(0.0051)
SC_share			-0.0003	-0.0003
			(0.0005)	(0.0005)
ST_share			-0.0000	-0.0000
			(0.0002)	(0.0002)
district_tsr			-0.0000	-0.0000
			(0.0000)	(0.0000)
avg_school_quality_index			0.0258	0.0348
			(0.0236)	(0.0257)
avg_household_characteristics_index			0.2421 ***	0.2504 ***
			(0.0255)	(0.0271)
avg_district_characteristics_index				-0.0311
				(0.0347)
R^2	0.2898	0.3480	0.5354	0.5363
Adj. R^2	0.2882	0.3450	0.5256	0.5254
Num. obs.	435	435	435	435

*** p < 0.001; ** p < 0.01; * p < 0.05

Model 1

$$\text{avg\_child\_edu} = \beta_0 + \beta_1 \times \text{avg\_maternal\_edu} + \varepsilon$$

Model 2

$$\text{avg\_child\_edu} = \beta_0 + \beta_1 \times \text{avg\_maternal\_edu} + \beta_2 \times \text{female\_lfp} + \varepsilon$$

Model 3

$$\text{avg\_child\_edu} = \underline{\text{Model 2}} + \text{controls} (\beta_3 \times \text{viirs\_annual\_mean} + \beta_4 \times \text{avg\_paternal\_edu} + \beta_5 \times \text{SC\_share} + \beta_6 \times \text{ST\_share} + \beta_7 \times \text{district\_tsr} + \beta_8 \times \text{avg\_school\_quality\_index} + \beta_9 \times \text{avg\_household\_characteristics\_index}) + \varepsilon$$

Model 4

$$\text{avg\_child\_edu} = \underline{\text{Model 3}} + \beta_{10} \times \text{avg\_district\_characteristics\_index} + \varepsilon$$

Model Progression Insights:

1. In Model 1, we find that a one-year increase in district-level maternal education is associated with a 4.7 percentage point rise in ASER scores, a relationship that is highly statistically significant ($p < 0.001$). This finding echoes Sunder et al's findings. (2019) on the importance of maternal influence in shaping home learning environments, and aligns with Afridi's research, which highlights the role of educated mothers in advocating for and supporting their children's education.
2. When we account for female agency in Model 2 by adding female labor force participation (LFP), it emerges as a significant and positive predictor of child learning outcomes (β = 0.0020***). This suggests that maternal economic empowerment complements educational investments, consistent with Afridi's argument that skilled female employment fosters aspirational parenting and greater educational engagement.
3. Interestingly, in Model 3, the coefficient on VIIRS nighttime lights our proxy for economic development and urbanity turns negative and remains significant (β = –0.0050***). This runs counter to expectations. A few possible explanations could work here: the presence of urban inequality, where the average development masks deep educational disadvantages in slums; migration-related instability in fast-urbanizing districts; or even measurement error, given the known limitations of nightlights in capturing equitable development at sub district levels.
4. Finally, in the fully adjusted Model 4, household characteristics emerge as the strongest predictor of ASER scores (β = 0.2504***), pointing to the foundational role of material well-being. Yet even after controlling for this, maternal education remains significant (β = 0.0315*)**, reaffirming that its contribution to child learning is not reducible to household wealth. Instead, it likely reflects a broader set of capabilities—ranging from literacy practices and time use to psychological investment in schooling—that mothers pass on to their children.
5. Across models, paternal education shows no significant effect on ASER outcomes, reinforcing the core finding in Sunder et al. (2019) that fathers' education has little bearing on children's early learning. Unlike maternal education, which translates into daily academic support and a richer home learning environment, paternal education appears too distal to matter. This result strengthens the consensus that maternal not paternal education is the true engine of foundational learning in the Indian context.

State Fixed Effects: Education Matters Within States Too:

	Model A
Maternal Education	0.0199 (0.0060) **
Female LFPR	0.0012 (0.0004) **
Night Time Lights	-0.0047 (0.0024) *
Paternal Education	0.0090 (0.0058)
SCs	-0.0004 (0.0005)
STs	-0.0003 (0.0002)
District TSR	0.0000 (0.0000)
School Quality Index	0.1450 (0.0351) ***
HH Characteristics Index	0.2296 (0.0359) ***
District Characteristics Index	-0.0644 (0.0408)
Num. obs.	435
R^2 (full model)	0.7047
R^2 (proj model)	0.3350
Adj. R^2 (full model)	0.6763
Adj. R^2 (proj model)	0.2712
Num. groups: state_name	29

*** p < 0.001; ** p < 0.01; * p < 0.05

Controlling for state-level heterogeneity such as administrative efficiency, or historical investment in schooling enables us to assess whether maternal education influences child outcomes within states, rather than just across states. After adding state fixed effects, the coefficient on maternal education drops to 0.0199* ($p = 0.001$), but it remains statistically significant. This attenuation is expected, as some of the initial correlation stemmed from high-education, high-outcome states like Kerala and Tamil Nadu. However, the persistence of the effect suggests that within any given state, districts with higher maternal education tend to have better ASER outcomes.

The female labor force participation (LFP) variable also remains significant (0.0012**), reinforcing Afridi's argument that empowered mothers are better able to engage with school systems, particularly in areas where state mechanisms may be weak. The household characteristics index (0.2296*) continues to be a strong predictor of outcomes, yet it does not overshadow the effect of maternal education, pointing to a potential social capital channel as discussed in Sunder et al. (2019). Additionally, the negative coefficient on VIIRS (-0.0047)* persists, suggesting that the issue is not confined to wealthier or poorer states but may be structural in nature—urban poverty and educational fragmentation could be embedded even within more developed regions. The adjusted R^2 increases to 0.6763, indicating a much better model fit. The robustness of maternal education's effect after accounting for state fixed effects strengthens its role as a within-state driver of educational outcomes.

Regional Heterogeneity: Not All Contexts Respond Equally

	North	South	East	West	Northeast	Central
Maternal Education	0.0253 (0.0095) **	0.0109 (0.0154)	0.0039 (0.0197)	0.0109 (0.0134)	0.0087 (0.0244)	0.0146 (0.0183)
Female LFPR	0.0030 (0.0006) ***	0.0009 (0.0011)	-0.0021 (0.0014)	0.0006 (0.0013)	-0.0003 (0.0016)	0.0032 (0.0014) *
Night Time Lights	-0.0039 (0.0032)	-0.0023 (0.0028)	-0.0189 (0.0162)	-0.0026 (0.0161)	0.0147 (0.0347)	0.0091 (0.0273)
Paternal Education	0.0285 (0.0108) **	-0.0125 (0.0123)	0.0321 (0.0195)	0.0114 (0.0124)	0.0304 (0.0207)	-0.0033 (0.0143)
SCs	0.0004 (0.0007)	-0.0023 (0.0011) *	0.0006 (0.0012)	0.0000 (0.0018)	0.0013 (0.0045)	-0.0053 (0.0031)
STs	0.0007 (0.0004)	-0.0008 (0.0013)	0.0003 (0.0009)	-0.0006 (0.0007)	0.0008 (0.0007)	-0.0023 (0.0009) *
District TSR	-0.0001 (0.0001)	-0.0026 (0.0009) **	0.0002 (0.0001) *	-0.0011 (0.0005) *	0.0004 (0.0002)	-0.0000 (0.0001)
School Quality Index	0.1028 (0.0608)	0.1793 (0.1101)	0.1365 (0.0791)	0.0467 (0.1304)	0.0905 (0.0842)	0.2144 (0.1089)
HH Characteristics Index	0.1742 (0.0562) **	0.5030 (0.1536) **	0.1626 (0.0973)	0.1763 (0.1098)	0.2689 (0.1108) *	0.0198 (0.1134)
District Characteristics Index	-0.0039 (0.0613)	-0.1932 (0.0791) *	0.2052 (0.1351)	0.0078 (0.1099)	-0.0425 (0.1166)	-0.0099 (0.1564)
Num. obs.	108	79	70	76	58	44
R^2 (full model)	0.9008	0.6540	0.6426	0.5850	0.6958	0.5605
R^2 (proj model)	0.6805	0.4655	0.5152	0.4564	0.4026	0.4014
Adj. R^2 (full model)	0.8846	0.5783	0.5596	0.4898	0.5770	0.4095
Adj. R^2 (proj model)	0.6284	0.3485	0.4026	0.3317	0.1695	0.1956
Num. groups: state_name	6	5	4	5	7	2

*** p < 0.001; ** p < 0.01; * p < 0.05

When the maternal education coefficient is allowed to vary by region, important regional patterns emerge. The effect is strongest in the North and Northeast. In the North, the coefficient is 0.0253**, remaining statistically significant at conventional levels. This mirrors Afridi's findings, suggesting that in more patriarchal northern settings, maternal education may play a particularly critical role in overcoming institutional gender biases and asserting children's rights in underperforming school systems. In the Northeast, the effect is marginally significant (0.0087), potentially reflecting the high cultural value placed on maternal roles and education.

In contrast, the effect of maternal education is weaker in the Central, South, East, and West regions. In the South, this may reflect a ceiling effect, as mothers are already more educated and schools are more functional, making it harder to detect marginal improvements. In Central and East India, the weak school systems may mean that maternal education alone cannot compensate for poor infrastructure and low parental capital, echoing Sunder et al.'s emphasis on "double disadvantage."

When examining complementary covariates by region, household characteristics are significant in the North, Central, and South regions, while the teacher-student ratio (TSR) is significantly negative in the South and East, where school staffing may vary more widely. The School Quality Index is mostly not significant, indicating limited variation or insufficient measurement. Model fit is highest in the North ($R^2 = 0.90$), where both explanatory variables and ASER outcomes vary sharply across districts. It is lowest in the West ($R^2 = 0.58$), possibly due to greater homogeneity in the region.

What This Tells Us

Maternal education appears to matter most in regions with weak schooling systems. In contexts like parts of the North, maternal education fills the gap through at-home instruction, pressure on local schools, and social capital networks. In contrast, in areas with stronger school systems (like the South), the marginal impact of maternal education shrinks because schooling itself may substitute for maternal involvement.

This supports Afridi's "interaction" view, where education is not a panacea but becomes particularly important when it enables women to address structural deficits. It also aligns with Sunder et al.'s notion of maternal "literacy as capability," where educated mothers are better able to leverage resources like school access to improve their children's learning outcomes.

6. Limitations

While our analysis offers some insights into the relationship between female human capital and child learning outcomes, it is important to situate these findings within the study's methodological limitations.

First, the cross-sectional nature of the data inherently restricts our ability to draw causal inferences. Although we have controlled for a broad range of potential confounding factors, the possibility of endogeneity arising from omitted variables or reverse causality cannot be fully ruled out. Therefore, the results should be read as evidence of associations, not as definitive proof of causality.

Second limitation lies in the spatial granularity of the ASER dataset. Though it is both nationally representative and methodologically robust, the ASER survey is designed to capture variation at the district level. It does not permit a closer look at disparities within districts, such as differences between villages or among urban neighborhoods. This lack of intra-district granularity limits the precision with which we can assess localized patterns of educational inequality.

The third limitation is about the issue of data missingness which also poses a challenge. A significant number of observations had to be dropped due to incomplete data, especially in key variables such as maternal education and household amenities. This reduction in the sample size (from 642 to 435 districts) not only decreases statistical power but also raises the possibility of a bias, which could influence the external validity of the findings.

Furthermore, the dataset offers only limited information on household-level socioeconomic conditions. In particular, the absence of direct income measures forces us to rely on a proxy index of household amenities to approximate economic status. While useful, this index may not fully capture the nuances of household wealth and consumption patterns.

Our measurement of female labor force participation also has notable constraints. The data, drawn solely from the Census, reduce women's economic activity to a binary employment variable. It offers no insight into the nature of the work performed, the number of hours worked, or participation in informal or unpaid labor- the dimensions that are crucial for a fuller understanding of women's economic roles.

Despite these limitations, the robustness of our regression models, especially the high R^2 value of approximately 0.74 in Model 4, indicates that the variables included explain a significant portion of the variation in child learning outcomes. This strengthens the argument that female human capital plays a central role in shaping educational attainment, even within the constraints of available data.

6. Conclusion

Our study affirms a powerful message: Female human capital matters immensely for child learning more so than urbanization, caste, or even paternal education.

This study provides new empirical evidence that maternal education and female labor force participation—two core dimensions of female human capital play a pivotal role in shaping children's foundational learning across Indian districts. Drawing on microdata from ASER 2011 and merging it with socio-economic indicators from SHRUG and the Census, we find a robust and consistent association between higher levels of maternal education and improved child learning outcomes, even after controlling for household assets, paternal education, caste, and school infrastructure. This relationship persists within states and is especially pronounced in regions where public schooling systems are weaker, suggesting that maternal education may act as a compensatory force where institutional capacity is limited.

The effects of female labor force participation, while more modest, remain statistically significant and positively associated with child learning—implying that beyond income effects, maternal economic agency may enhance educational investments and advocacy within the household. Interestingly, traditional proxies of development such as night-time light intensity fail to predict better learning outcomes and, in some specifications, show a negative association—highlighting the limits of urbanization as a standalone driver of human development.

Several implications emerge from these findings. First, maternal education is not merely a background characteristic but appears to function as an active lever for intergenerational mobility, particularly in resource-constrained settings. Second, economic empowerment of women—through work opportunities or targeted schemes like MGNREGA—may yield educational externalities for the next generation. Third, material household conditions significantly shape the transmission of human capita - our household characteristics index consistently shows the largest effect size, underlining the intertwined nature of gender and economic inequality.

Nevertheless, these results must be interpreted with caution. The ecological and cross-sectional design limits causal claims, and district-level aggregation obscures within-district inequalities and heterogeneity. Moreover, our measures of female labor force participation do not account for informal or unpaid labor, which remain dominant in India's gendered economy.

Future research would benefit from longitudinal data to capture dynamic changes in maternal roles and their effects, as well as mixed-methods designs that uncover the lived experiences behind these associations. As India continues to grapple with persistent educational inequality, this study reaffirms that enhancing female human capital is not only a question of gender justice—but one of educational justice as well.

7. Next Steps

While this study provides robust district-level evidence on the relationship between female human capital and child learning, it opens several avenues for deeper empirical and theoretical engagement.

1. Move Toward Causal Inference.
Future research should employ quasi-experimental strategies to identify causal effects of maternal education and labor force participation. For instance, phased rollouts of programs like MGNREGA or historical variations in education policy (e.g., DPEP expansion) could serve as instruments to address endogeneity. Establishing whether maternal education *causes* improved child outcomes—rather than merely co-occurring with them—is essential for designing effective interventions.

2. Leverage Longitudinal and Panel Data.
Analyzing change over time using multiple rounds of ASER or panel surveys like Young Lives could illuminate how shifts in maternal characteristics and household conditions shape learning trajectories. In particular, the long-awaited release of the **next Population Census** will offer critical opportunities to update district-level indicators such as FLFP and demographic composition—allowing for replication, trend analysis, and dynamic modeling beyond the 2011 baseline.

3. Disaggregate and Deepen the Analysis.

Disaggregating child learning outcomes by gender, age cohort, or caste group could reveal important intersectional dynamics that are masked in aggregate models. Moreover, distinguishing between types of maternal work—formal, informal, unpaid—would provide a more nuanced understanding of how economic participation affects educational investment.

4. Complement Quantitative Findings with Qualitative Insight.

Quantitative analysis alone cannot fully uncover the behavioral and cultural mechanisms at play. Ethnographic studies or in-depth interviews in contrasting districts could explore how maternal literacy translates into educational support, how work reshapes intra-household priorities, and how women navigate school systems on behalf of their children.

Together, these next steps would not only strengthen the empirical foundations of this research but also enhance its relevance for both policy and practice. Understanding how, when, and why maternal capabilities shape children's futures is critical for any serious attempt to address India's persistent learning inequalities.

Works Cited

1. "Annual Status of Education Report (Rural) 2018." ASER Centre, 2018. https://asercentre.org/aser-2018/RISE Programme+2ASER: Annual Status of Education Report+2img.asercentre.org+2

2. Banerjee, R., Barry, J., & Shortland, M. (n.d.). Asercentre. https://img.asercentre.org/docs/Publications/Other%20Publications/maternalliteracy_rukminibanerji.pdf

3. Doctor, T. (2024, November 5). Impact of maternal employment on child academic performance aged 16 - 20 years an Indian perspective. SCIRP. https://www.scirp.org/journal/paperinformation?paperid=137782
4. Sunder, N. (2020). Parents' schooling and intergenerational human capital: Evidence from india.

5. Afridi, Farzana, Abhiroop Mukhopadhyay, and Soham Sahoo. "Female Labor Force Participation and Child Education in India: Evidence from the National Rural Employment Guarantee Scheme." *IZA Journal of Labor & Development*, vol. 5, no. 1, 2016, pp. 1–27. https://izajold.springeropen.com/articles/10.1186/s40175-016-0053-yIZA+4SpringerOpen+4IDEAS/RePEc+4

6. Luke, Nancy, and Kaivan Munshi. "Women as Agents of Change: Female Income and Mobility in India." *Journal of Development Economics*, vol. 94, no. 1, 2011, pp. 1–17. https://ideas.repec.org/a/eee/deveco/v94y2011i1p1-17.htmlResearchGate+6IDEAS/RePEc+6EconPapers+6

7. Qian, Nancy. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *The Quarterly Journal of Economics*, vol. 123, no. 3, 2008, pp. 1251–1285. https://academic.oup.com/qje/article-abstract/123/3/1251/1928174Kellogg School of Management+7Oxford Academic+7ResearchGate+7

8. Sahoo, Soham, and Ranjan Kumar. "The Impact of MGNREGS on Child Labour and Child Education: An Empirical Analysis." *Indian Journal of Labour Economics*, vol. 59, no. 3, 2016, pp. 407–425. https://www.researchgate.net/publication/329812402_The_impact_of_MGNREGS_on_c

hild_labour_and_child_education_an_empirical_analysisResearchGate

9. Sunder, Nandini. "Education as Empowerment: The Role of Education in the Empowerment of Marginalized Communities in India." *Economic and Political Weekly*, vol. 54, no. 4, 2019, pp. 45–53.

10. Vikram, K., Chen, F., & Desai, S. (2018, February 16). Mothers' work patterns and children's cognitive achievement: Evidence from the India Human Development Survey. Social science research. https://pmc.ncbi.nlm.nih.gov/articles/PMC6003423

11. SHRUG: @misc{pcindia, title={Census of India}, author={Office of the Registrar General and Census Commissioner, India}, year={2011}, note={Accessed via SHRUG on May 08, 2025 \url{https://www.devdatalab.org/shrug_download/}, v2.1.pakora}

Appendix 1:

Table 1: State Wise Regional Distribution

Region	**States**
East	Bihar, Jharkhand, Odisha, West Bengal
West	Rajasthan, Gujarat, Maharashtra, Goa, Dadra Nagar Haveli Daman Diu, Daman Diu
North	Jammu Kashmir, Himachal Pradesh, Punjab, Uttarakhand, Haryana, Delhi, Uttar Pradesh, Chandigarh
South	Andhra Pradesh, Telangana, Karnataka, Kerala, Tamil Nadu, Puducherry
Central	Chhattisgarh, Madhya Pradesh
North East	Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Meghalaya, Assam