How does literacy rate affects country's GDP per capita?

Aldiyar Rustem

Nazarbayev University

April 28, 2025

1 Introduction

For many years economists together with public policymakers have acknowledged the existence of an essential correlation between national literacy levels and economic success. Human civilization relies critically on literacy rates because these factors power the development of the labour market and both innovative power and technological adjustments leading to economic expansion. The worldwide growth of literacy rates shows uneven development because literacy levels remain high in particular regions of specific countries. A total of 750 million adults lack basic reading and writing capabilities and this deficiency is mostly prevalent in the regions of sub-Saharan Africa and South Asia (UNESCO, 2020). The different levels of literacy show the disparity caused by economic factors which influence national economic achievements. The GDP value per capita of the countries will serve as our evidence to understand how literacy affects national economic effectiveness.

A study analyzes how adult literacy rates influence GDP per capita through nation-based data collection efforts to verify the role that education plays in economic development. The current research focuses on this matter by addressing Sustainable Development Goals such as SDG 4 (Quality Education) and SDG 8 (Decent Work and Economic Growth). This research controls macroeconomic elements and population characteristics to examine how literacy rates affect GDP per capita so as to assist policy establishment for education-driven sustainable development.

2 Data Description

A reputable public database from The World Bank (2016) provides standardized economic and social indicators. The dataset was retrieved from different social-economic indicators available there and combined into one. It includes multiple variables as follows:

- **Dependent Variable:**
  - **GDP per capita (US$)**: Measures economic prosperity in terms of the volume of production in a national economy per population. It is the main indicator of economic success.
- **Independent Variables:**
  - **Adult Literacy Rate (% ages 15+)**: Basic reading and writing skills levels among people aged 15 and older. The variable measures human capital influence on productivity in a direct manner.
  - **Inflation (%)**: Annual market price changes, which measures the influence

of economic stability on growth.

- **Urban Population (%)**: Shows the influence of urban population growth on industrial development and labor productivity.
- **Exports (% of GDP)**: The variable measures trade openness because this factor is a known driver of economic growth.
- **Labor Force Participation (%)**: Measures the proportion of the population who are economically active.
- **Unemployment (%)**: Measures the percentage of the labor force who are unemployed.
- **Government Education Expenditure (% of GDP)**: Reflects government investment in human capital.
- **Primary Completion Rate (%)**: Measures how well educated people are, after primary school completion.

Variable	Obs	Mean	Std. dev.	Min	Max
GDP per Capita	256	15 909.1	23 750.5	235.3	173 604.8
Literacy Rate	80	82.1	16.3	22.0	100.0
Urbanization Rate	261	59.2	22.9	12.4	100.0
Exports	228	38.4	28.3	3.2	191.1
Inflation Rate	234	6.2	30.2	–3.1	380.0
Unemployment Rate	211	7.6	5.4	0.2	26.6
Labor Force Participation Rate	148	61.3	9.1	34.1	91.8
Government Spending on Educ.	110	4.5	1.5	1.4	8.1
Primary Completion Rate	179	91.1	14.4	40.9	131.0
log(GDP per Capita)	256	8.8	1.4	5.5	12.1

Table 1: Summary Statistics of Each Variable

3 Literature review

- Hanushek and Woessmann (2008): The research paper shows that literacy and numeracy are key elements in determining economic growth rates over extended periods. The analysis of data from the 1960-2000 period indicates that every additional one standard deviation in cognitive skills can increase annual GDP growth rate by 2%. The researchers establish that general population literacy upgrades lead to superior economic growth than specialized schooling does because national productivity depends on average skill levels. According to their research the economic development of both individual workers and national economies depends on literacy skills.

- Becker and Woessmann (2009): In their seminal work, the authors oppose the Protestant idea of hard work leading to economic growth. They demonstrate through Prussian 19th-century data how Protestant areas achieved better economic results than Catholic regions due to initiatives on compulsory Bible reading that contributed to the development of reading skills. Using the Wittenberg distance (the origin of Protestantism) as an instrumental variable has helped the researchers to identify how literacy created its causal impact. They found that Protestant reading abilities explained almost everything in GDP differences between areas.

The two analyses demonstrate how important reading skills are for achieving economic advancement. The evidence provided by Hanushek and Woessmann (2008) reveals a contemporary worldwide model yet Becker and Woessmann (2009) deliver historical evidence about literacy's causal effect. The findings from both studies confirm that literacy can significantly influence the variations in GDP per capita.

4 Methodology

4.1 Econometric Models

We estimate four specifications of the relationship between literacy rates and GDP per capita using Ordinary Least Squares (OLS). Each model introduces additional controls to assess robustness and address confounding factors.

- **Model 1: Baseline Specification**

$$\log(GDP_i) = \beta_0 + \beta_1 Literacy\ Rate_i + u_i$$

  - **Purpose**: Estimate the unconditional effect of literacy on GDP.
  - **Variables**: Literacy rate (sole predictor).

- **Model 2: Core Controls**

$$\log(GDP_i) = \beta_0 + \beta_1 Literacy\ Rate_i + \beta_2 Inflation_i + \beta_3 Urbanization_i + \beta_4 Exports_i + u_i$$

  - **Purpose**: Isolate literacy's effect while controlling for macroeconomic and structural factors.
  - **Added Variables**: Inflation, urbanization, exports.

- **Model 3: Labor Market and Policy Controls**

$$\log\Big(\mathrm{GDP}_i\Big) = \beta_0 + \beta_1\,\mathrm{LiteracyRate}_i + \beta_2\,\mathrm{Inflation}_i + \beta_3\,\mathrm{Urbanization}_i + \beta_4\,\mathrm{Exports}_i + \beta_5\,\mathrm{Unemployment}_i + \beta_6\,\mathrm{GovEducationSpending}_i + u_i$$

  - **Purpose**: Account for labor market dynamics and policy investments.
  - **Added Variables**: Unemployment rate, government education spending.

- **Model 4: Extended Human Capital Specification**

$$\log\Big(\mathrm{GDP}_i\Big) = \beta_0 + \beta_1\,\mathrm{LiteracyRate}_i + \beta_2\,\mathrm{Inflation}_i + \beta_3\,\mathrm{Urbanization}_i + \beta_4\,\mathrm{Exports}_i + \beta_5\,\mathrm{GovEducationSpending}_i + \beta_6\,\mathrm{LaborParticipationRate}_i + \beta_7\,\mathrm{PrimaryCompletionRate}_i + u_i$$

  - **Purpose**: Test robustness with alternative human capital metrics.
  - **Added Variables**: Labor force participation rate, primary completion rate.
  - **Limitation**: Severe missingness reduces sample size to 20 observations.

5 Results

Variable	Model 1	Model 2	Model 3	Model 4
Literacy Rate	0.051***	0.024***	0.019*	0.003
	(0.005)	(0.004)	(0.009)	(0.025)
Urbanization		0.024***	0.024***	0.021**
		(0.004)	(0.006)	(0.008)
Inflation		−0.039*	−0.001	−0.015
		(0.020)	(0.033)	(0.042)
Export		0.012***	0.025**	0.023**
		(0.003)	(0.008)	(0.009)
Unemployment			−0.001	0.003
			(0.019)	(0.037)
Gov. Education Spend.			0.100	0.003
			(0.084)	(0.151)
Labor Rate				0.000
				(0.025)
Primary Rate				0.019
				(0.028)
Cons	4.082***	4.692***	4.185***	4.503*
	(0.399)	(0.298)	(0.691)	(2.341)
R-squared	0.595	0.829	0.810	0.7265
F	113.30***	84.91***	16.35***	3.652*
Obs	79	75	30	20

Table 2: Regression Results: Economic Indicators (The World Bank, 2016). $^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$. Standard errors are in parentheses.

5.1 Preferred Model

The regression results demonstrate Model 2's dominance as our primary specification. It almost perfectly explains the variation ($R^2 = 0.829$) in GDP per capita, while also does not suffer from sample size reduction as well as controls for the main macroeconomic variables. While Model 1 shows the strongest effect (0.051% GDP increase per 1% literacy gain) of literacy rate on economic prosperity, its simplicity introduces omitted variable bias. Models 3 and 4 should be rejected as the critically small sample sizes (n=30 and n=20 respectively) do not fully illustrate the literacy rate's effect on our main independent variable and undermine the coefficients. Overall, Model 2 has to be accepted as the primary preferred model as it not only maintains a statistically significant effect

of the literacy rate (0.024%, p <0.01) but also controls for important economic drivers, urbanization rate (0.024%) and exports (0.012%), (unlike Model 1) and does not suffer from sample size reduction (unlike Models 3 and 4).

5.2 Heteroskedasticity

To test for heteroskedasticity—non-constant variance in the error terms—we employ the White test, which generalizes the Breusch-Pagan test by including squared terms and interactions of predictors. The test is implemented as follows:

$$\hat{u}_i^2 = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + \gamma_3 X_i X_j + \cdots + \epsilon_i$$

For our primary specification (Model 2), the White test yields:

- **p-value**: 0.737

Interpretation: The large p-value (> 0.05) fails to reject the null hypothesis, indicating no evidence of heteroskedasticity. This validates the use of standard OLS inference. There is no need to use robust standard errors as they can unmeaningfully harm our model.

6 Limitations

1. **Missing Data and Sample Size Constraints**: Since the literacy rate variable we are interested in has a lot of missing observations we are limited by our analysis. In our primary specification (Model 2), we have 188 missing values (71% of the 263-country original dataset), therefore, our final sample size is 75 observations. This raises three concerns:

    - **Reduced Statistical Power**: Smaller samples increase standard errors by making it more difficult to view true effects of our results.
    - **Selection Bias**: Because the included countries with literacy data may differ systematically from those with missing literacy data (often, low income nations), generalization of the findings may be limited.
    - **Model Instability**: As additional controls are added (e.g., unemployment, education spending) sample size is reduced to 20 observations in Model 4 and hence coefficients become unreliable.

2. **Measurement Validity Trade-offs**: To address missingness, we considered substituting literacy rate with primary completion rate (84 NAs).

   Completing primary school provides access to education for children, while literacy tests adults.

   Currently received primary school completion will affect adult literacy rates after a time delay of 10–15 years. However this does not align with the data from 2016 GDP data.

   Primary school completion does not ensure students become literate.

3. **Interpretation of Limitations**.

   High–literate nations are better at collecting and reporting accurate data, so we get more figures from them when estimating literacy rates.

   Literacy rates are accurate but not always available, and primary completion rates are available but not as accurate as literacy rates.

7 Conclusion

7.1 Main Conclusion

The paper evaluates how literacy rates affect GDP per capita using worldwide databases from The World Bank (2016). The analysis uses four regression models to add sequential controls for different macroeconomic and structural and policy elements. Three important findings:

- Literacy rates show a strong correlation with GDP per person results in the basic models which indicate that every 1% increase in literacy leads to a 0.024% to 0.051% increase in GDP. Model 4 includes several additional potentially unreliable variables which produces a minimal and nonsignificant impact on the effect. The research proves that both incomplete data and advanced model structures influence the study findings.
- Urbanization ($\beta \approx 0.024$) together with exports ($\beta \approx 0.025$) serve as reliable variables because both city living and international trade contribute positively to economic results.
- The analysis suffers from a significant lack of data on literacy rates because 69% of the dataset entries are missing. The large amount of missing figures in literacy rates creates reduced statistical strength and decreased model generalization, especially

when working with complex models

7.2 Policy Implications

The findings of the study are consistent with human capital theory and evidence from history (Hanushek & Woessmann, 2008) furthermore support policies that focus on:

- **Adult literacy programs** designed to improve basic skills among unprivileged populations.
- **Extra government funding** for primary education will help resolve intergenerational literacy gaps.
- **Infrastructure development** that creates a connection between urban growth with export capabilities

7.3 Limitations and Future Research

Three analytical limitations must be considered in the study's empirical investigation.

- Cross-sectional data preclude causal claims. It only tells about correlation, not necessarily causation.
- Literacy rate statistics may fail to measure the level of functional skills. Validity could be enhanced by complementing with PIAAC style assessments.
- The problem of missing data leads to excessive exclusion of countries that maintain lower levels of income. The evaluation would be improved by using imputation techniques or alternative datasets such as those from UNESCO (2020).

Economic development shows a direct correlation to literacy rate, but this relationship depends on various environmental factors as well.

8 References

Becker, S. O., & Woessmann, L. (2009). Was weber wrong? a human capital theory of protestant economic history. *The quarterly journal of economics*, *124*(2), 531–596.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of economic literature*, *46*(3), 607–668.

The World Bank. (2016). Economic indicators [Data retrieved from World Development Indicators, https://data.worldbank.org/indicator].

UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education-all means all*. UN.

A Additional Results

Variable	Coef.	Std. Err.	t	$P > \|t\|$	[95% CI]
Literacy Rate	0.051***	0.0048	10.64	0.000	[0.0413, 0.0603]
Constant	4.082***	0.3986	10.24	0.000	[3.2886, 4.8760]
Observations = 79		R-squared = 0.595		F-statistic = 113.27***	

Table 3: Model 1: Simple Linear Regression

Variable	Coef.	Std. Err.	t	$P > \|t\|$	[95% CI]
Literacy Rate	0.024***	0.0045	5.46	0.000	[0.0155, 0.0334]
Urbanization	0.024***	0.0038	6.44	0.000	[0.0167, 0.0318]
Exports	0.012***	0.0028	4.43	0.000	[0.0068, 0.0181]
Inflation	−0.039*	0.0198	−2.00	0.050	[−0.0789, 0.0000]
Constant	4.692***	0.2983	15.73	0.000	[4.0975, 5.2873]
Observations = 75		R-squared = 0.829		F-statistic = 84.91***	

Table 4: Model 2: 4-Variable Multiple Linear Regression

Variable	Coef.	Std. Err.	t	$P > \|t\|$	[95% CI]
Literacy Rate	0.019*	0.0095	2.02	0.055	[−0.0004, 0.0388]
Urbanization	0.024***	0.0060	3.93	0.001	[0.0112,0.0360]
Exports	0.025***	0.0080	3.08	0.005	[0.0081,0.0411]
Inflation	−0.001	0.0326	−0.04	0.967	[−0.0688, 0.0660]
Unemployment	−0.001	0.0192	−0.06	0.956	[−0.0407, 0.0385]
Gov. Spending	0.100	0.0837	1.19	0.246	[−0.0735, 0.2729]
Constant	4.185***	0.6905	6.06	0.000	[2.7569, 5.6138]
Observations = 30		R-squared = 0.810		F-statistic = 16.35***	

Table 5: Model 3: 6-Variable Multiple Linear Regression

Variable	Coef.	Std. Err.	t	$P > \|t\|$	[95% CI]
Literacy Rate	0.003	0.0250	0.12	0.909	$[-0.0520, 0.0579]$
Urbanization	0.021**	0.0077	2.71	0.020	[0.0039,0.0380]
Exports	0.023**	0.0092	2.50	0.030	[0.0027,0.0430]
Inflation	-0.015	0.0424	-0.35	0.736	$[-0.1080, 0.0787]$
Unemployment	0.003	0.0367	0.09	0.929	$[-0.0774, 0.0840]$
Gov. Spending	0.003	0.1506	0.02	0.983	$[-0.3282, 0.3348]$
Labor Rate	0.000	0.0247	0.01	0.995	$[-0.0542, 0.0545]$
Primary Rate	0.019	0.0282	0.67	0.515	$[-0.0430, 0.0810]$
Constant	4.503*	2.3410	1.92	0.081	$[-0.6499, 9.6550]$
Observations = 20		R-squared = 0.727		F-statistic = 3.65*	

Table 6: Model 4: Fully Specified Multiple Linear Regression

Test	χ^2	df	p-value
White's Test	11.70	14	0.630
Cameron & Trivedi's Decomposition:			
Heteroskedasticity	11.70	14	0.630
Skewness	3.09	4	0.543
Kurtosis	1.31	1	0.253
Total	16.09	19	0.651

Table 7: White's Test for Heteroskedasticity