Journal of Education and Educational Policy Studies Vol.3 Issue 3 EWA Publishing
Available Online: 15 July 2025 DOI: 10.54254/3049-7248/2025.24686

Study on influencing factors of education completion rate: a research on the relationship between school construction and different levels of education based on linear regression

Chenbing Zhou

Department of Economics, University of Colorado Denver, Denver, USA

chenbing0402@outlook.com

Abstract. As global focus on educational investment grows, the link between school infrastructure and education quality remains underexplored. While prior studies emphasize tuition fees and exam systems, the role of physical infrastructure investment lacks empirical evidence. This paper examines how government education spending, compulsory education duration, teacher qualifications, and enrollment rates interact across educational levels, offering policymakers insights to optimize investments and advance sustainable education goals.Using UNESCO and World Bank data (1970–2023), we analyze national education expenditure, compulsory education years, enrollment rates, and income levels via linear regression. Key variables include education expenditure as a percentage of GDP, compulsory education duration, income, qualified teacher ratios, and school internet access. Findings reveal that higher government education spending strongly correlates with increased primary/secondary enrollment, especially in low-income countries. Extending compulsory education boosts enrollment in low- and middle-income nations. Teacher qualifications significantly drive higher education participation, while internet access effects vary regionally.Investments in teacher training and school infrastructure are pivotal for improving enrollment. Policymakers should prioritize optimizing education budgets and resource allocation to enhance equity and quality.

Keywords:education completion rate, linear regression, school construction, education levels

1. Introduction

Governments around the world have long recognized the necessity of investing in education, as education serves as the cornerstone of social and economic progress. Among the many factors influencing educational outcomes, government school construction plays a vital role in shaping educational opportunities and quality. Factors such as the proportion of education expenditure in government GDP, the duration of compulsory education, income level, the proportion of qualified teachers, and the proportion of schools with internet access all have significant impacts on educational outcomes. The significance of our research lies in its ability to inform academic discussions and policymaking regarding educational advancement.

Despite the attention given to factors such as tuition costs and examination systems, little is known about the direct impact of physical infrastructure investment on educational participation. Therefore, our study fills this gap by exploring the relationship between government school construction and enrollment rates at different educational stages. Understanding these relationships is crucial for policymakers to design targeted interventions, maximize the return on government investment in education, and ultimately contribute to achieving sustainable development goals in education.

2. Literature review

The research by Max Roser and Esteban Ortiz Ospina (Roser M. & Ortiz-Ospina E., 2016) elaborates on the history, development, and other aspects of education financing, detailing methods of education financing and global trends in education expenditure, while conducting an in-depth analysis of various factors determining education. However, this article does not further explore the relationship between government education expenditure and educational attainment. Our study will further discuss the role of the proportion of education expenditure in government GDP in school construction and how it affects enrollment rates.

Additionally, some studies have explored the impact of government compulsory education duration on enrollment rates in small-scale settings. Luis Diaz Serrano et al. analyzed the impact of primary compulsory education duration on secondary education enrollment rates using panel data from non-OECD countries from 1970 to 2012 (Díaz Serrano, L., & Pérez Reynosa, J. H., 2018). The study shows that in low-income countries, extending compulsory education has high economic value for children, but parents have low willingness to let their children continue their education. This study uses more data to explore not only low-income countries but also the impact of compulsory education years on enrollment rates in high-income countries, providing a broader and more comprehensive reference for studying the impact of government construction on enrollment rates.

Some studies show that in some developed countries, parents' income is positively correlated with their children's educational attainment. However, Katrine V. Løken's Family Income and Children's Education: A Natural Experiment of the Norwegian Oil Boom* (Løken, K. V., 2010) presents a relatively different perspective. By analyzing a unique dataset of Norwegians born between 1967 and 1969, which measures the permanent family income of their children during adolescence, the study examines the long-term impact of family income on children's educational level. However, the context of this study is the Norwegian oil crisis in the 1970s, which meant a general increase in income. The study found no clear causal relationship between family income and children's educational level in Norway. In the study by Stephens et al. (Stephens Jr, M., & Yang, D. Y., 2014), various factors including compulsory education policies and laws are effectively explained as related to educational attainment. However, the data collection in this study is limited and more偏向于 the white population in U.S. states, which may impose certain limitations on the research. Our study will draw on broader data to explore the relationship between income level and educational level, in order to assess the impact of government investment in school construction on enrollment rates and obtain more general conclusions.

Olubunmi Kayode Aya Nwoye's article mainly discusses the impact of teacher development on educational attainment in Nigeria (Olubunmi Kayode Ayanwoye., 2023). The article points out that improving the quality of the teaching staff is equivalent to improving student performance. However, this study does not mention the relationship between the proportion of qualified teachers at different educational stages and enrollment rates. In our study, we will further investigate this issue and indicate the impact of the proportion of qualified teachers on government school construction, further exploring how this variable affects enrollment rates.

In Stephens' study (Stephens Jr, M., & Yang, D. Y., 2014), various factors including compulsory education policies and laws are effectively explained as related to educational attainment. However, the data collection in this study is limited and more prefer to the white population in U.S. states, which may impose certain limitations on the research.

Some studies have also analyzed the proportion of schools with internet access for teaching purposes and shown significant differences across regions, but they do not analyze their specific relationship with educational outcomes. Benchea's study specifically examines the dynamics of internet use in Romania and its impact on students' academic engagement (Benchea, L., 2021), concluding that there is a negative correlation between academic engagement and internet use. However, this article does not specifically focus on the digital divide between regions and its impact on educational attainment, in order to better determine the impact of government investment in school construction on enrollment rates.

3. Theoretical analysis

3.1. Human capital theory

The human capital theory posits that education and training are investments in human resources, which enhance individual productivity and economic value by improving skills and knowledge. In this paper, the impact of education expenditure and compulsory education duration on enrollment rates can be explained by the human capital theory. Increasing government education investment and extending compulsory education duration will improve individuals' educational levels, thereby enhancing the quality of the labor force in the labor market and promoting economic development. According to the human capital theory, more education investment helps improve individuals' professional skills, knowledge reserves, and innovative capabilities, ultimately increasing their employment opportunities and income levels. This theory emphasizes the importance of education for economic development and social progress and supports policies to extend compulsory education years and increase education expenditure, arguing that these measures can effectively enhance a nation's overall competitiveness.

3.2. Education equality theory

The education equality theory emphasizes that fairness in educational opportunities is an important embodiment of social justice. The theory holds that everyone, regardless of social background or economic conditions, should have equal access to education. In this paper, extending compulsory education years and increasing education expenditure help promote educational equity, ensuring that children from different social strata and regions can receive basic education. This theory supports the government in bridging gaps in educational resources through policy measures, especially in poor areas and among vulnerable groups.

According to the education equality theory, education should not be attachment of wealth and social status, and all citizens should have the opportunity to achieve social mobility through education. This theory helps explain why government education expenditure and the extension of compulsory education years can effectively improve enrollment rates, particularly for low-income families and students in remote areas, as providing more educational opportunities can break the intergenerational transmission of poverty.

4. Model hypotheses

In this paper, regression models are constructed for the gross enrollment rates at different educational levels. The dependent variables include: y1 (prienroll, Primary school enrollment rate), y2 (secenroll, Secondary school enrollment rate), y3(terenroll, Tertiary education enrollment rate).The explanatory variables are: x1 (spen, Government education expenditure as % of GDP), x2 (eduyear, Duration of compulsory education in years), x3 (highincome, High-income country dummy: 1=yes, 0=no),x4 (teacher, Percentage of qualified teachers),x5 (internet, Proportion of schools with internet access).

Hypothesis 1 (H1): Government education expenditure (x1) has a significant positive impact on enrollment rates.

Primary education (y1): Coefficient is negative (β=-0.0006) but not significant ($p>0.05$), hypothesis not supported.

Secondary education (y2): Coefficient is significantly positive (β=0.0215, $p<0.05$), hypothesis supported.

Higher education (y3): Coefficient is significantly positive (β=0.0656, $p<0.01$), hypothesis supported.

The positive effect of education expenditure increases with educational level, possibly due to unequal resource allocation or structural issues weakening its effect at the primary level.

Hypothesis 2 (H2): Compulsory education duration (x2) has a significant positive impact on enrollment rates.

Primary education (y1): Coefficient is negative (β=-0.0011) and not significant ($p>0.05$), hypothesis not supported.

Secondary education (y2): Coefficient is significantly positive (β=0.0282, $p<0.01$), hypothesis supported.

Higher education (y3): Coefficient is significantly positive (β=0.0351, $p<0.05$), hypothesis supported.

Compulsory education policies have a more pronounced effect on secondary and higher education enrollment rates, with primary-level effects possibly offset by policy implementation differences (e.g., dropout rates).

Hypothesis 3 (H3): High-income countries (x3=1) have a significant positive impact on enrollment rates.

Primary education (y1): Coefficient is negative (β=-0.0082) and not significant ($p>0.05$), hypothesis not supported.

Secondary education (y2): Coefficient is significantly positive (β=0.1281, $p<0.01$), hypothesis supported.

Higher education (y3): Coefficient is significantly positive (β=0.5061, $p<0.01$), hypothesis supported.

Economic level has a much stronger impact on secondary and higher education than primary education, possibly due to high-income countries' greater emphasis on secondary and higher education resource investment.

Hypothesis 4 (H4): Proportion of qualified teachers (x4) has a significant positive impact on enrollment rates.

Primary education (y1): Coefficient is significantly negative (β=-0.0010, $p<0.05$), hypothesis not supported and direction opposite.

Secondary education (y2): Coefficient is positive but not significant (β=0.0008, $p>0.05$), hypothesis not supported.

Higher education (y3): Coefficient is positive but not significant (β=0.0041, $p>0.05$), hypothesis not supported.

The unexpected result at the primary level may be due to mismatched teacher resources (e.g., "qualified but inefficient"), while at the secondary and higher levels, teacher quality may not be fully measured (e.g., research capabilities not included in indicators).

Hypothesis 5 (H5): Proportion of schools with internet access (x5) has a significant positive impact on enrollment rates.

Primary education (y1): Coefficient is negative (β=-0.0001) and not significant ($p>0.05$), hypothesis not supported.

Secondary education (y2): Coefficient is significantly positive (β=0.0031, $p<0.05$), hypothesis supported.

Higher education (y3): Coefficient is significantly positive (β=0.0077, $p<0.01$), hypothesis supported.

Internet facilities are more attractive for secondary and higher education, likely due to the higher value of digital teaching tools in complex curricula.

5. Empirical analysis

5.1. Data source

Our study uses cross-sectional data from the UNESCO Institute for Statistics and the World Bank. These data include information on government education expenditure, compulsory education years, school enrollment rates, household income levels, etc., from 1970 to 2023.

The data used in this study come from two authoritative sources: the UNESCO Institute for Statistics and the World Bank. These cross-sectional datasets provide comprehensive information on key educational and socioeconomic indicators, including government education expenditure, compulsory education duration, enrollment rates, and household income levels. Spanning a

wide time range from 1970 to 2023, these datasets offer valuable insights for studying long-term trends and dynamic interactions between education investment and socioeconomic factors. The inclusion of macroeconomic indicators and individual education statistics allows for a detailed examination of the relationships between government policies, household income, and educational outcomes across different national contexts.

The first step was to load five separate datasets into Stata software, each containing country and year information. These datasets included government investment data, school enrollment data, and household income data, among others. The integrity and consistency of variable names and formats in each dataset were checked. The datasets were then merged one by one using Stata's merge command, with country and year variables as common identifiers. This completed the preliminary data analysis.

5.2. Model

In this paper, regression models are constructed for the gross enrollment rates at different educational levels. The proportion of government education expenditure in GDP is selected as the main explanatory variable x1, while compulsory education duration, national income classification, proportion of qualified teachers, and the proportion of schools with internet access for teaching purposes are used as other explanatory variables x2-x5 to jointly explain the dependent variables y1, y2, and y3. The meanings of variables in the model are shown in Table 1.

This model explores the relationship between gross enrollment rates at different levels and school construction primarily measured by government education expenditure. National investment in education is a crucial component of school construction. Other variables, including compulsory education duration, national income classification, proportion of qualified teachers, and the proportion of schools with internet access for teaching purposes, all influence gross enrollment rates and measure school construction, which is relevant to our research question.

5.3. Descriptive statistics

Variables cover aspects such as educational level (primary, secondary, higher education), education expenditure, compulsory education duration, national income level, proportion of qualified teachers, and proportion of schools with internet access. Below is detailed information on each variable.

Table 1 and Table 2 illustrate the distribution of the percentage of qualified teachers, showing a right-skewed pattern that implies teacher shortages in some countries.

[Table 1 and table 2 here]

Primary school enrollment rate (y1). This variable reflects the proportion of students within the school-age range (typically 6-12 years old) receiving education. The average value is 102.6, indicating that the proportion is slightly above 100% on average, meaning that in some regions, the number of students enrolled in this age group may be higher than the expected 100%. This is due to double counting. For example, in areas with high population mobility, some students may be repeatedly counted in enrollment statistics across different regions. Its standard deviation indicates some variation across countries, but the variation is not large. At the primary education level, most countries have high enrollment rates, which generally signifies relatively sound education policies and resource allocation in these regions.

Secondary school enrollment rate (y2). This variable represents enrollment in junior and senior high school education, usually expressed as the ratio of junior and senior high school students to the total population in that age group. The average value is 92.19, indicating that approximately 92% of students are enrolled in junior or senior high school on average, slightly lower than the primary school enrollment rate. The larger standard deviation indicates significant variation in enrollment rates across countries. The increase in secondary school enrollment rates generally reflects the expansion of education systems and the success of government education policies.

Higher education enrollment rate (y3). The higher education enrollment rate refers to the level of participation in university and above-level education, usually expressed as the ratio of students in the university-age group to the population in that age group. The average value is 47.16, indicating that approximately 47% of students participate in university or higher-level education on average, far lower than secondary school enrollment rates. This suggests that higher education is not given sufficient attention in many countries. The large gap between the minimum and maximum values reveals significant regional disparities.

Proportion of education expenditure in total government expenditure (x1). This variable refers to the government's financial investment in education, including expenditures by governments at all levels on school construction, teacher salaries, educational resources, and curriculum development. On average, the proportion of education expenditure in GDP is 4.279%. Different countries and regions have varying degrees of emphasis on education and economic strength, leading to differences in the proportion of education expenditure. A smaller standard deviation means that the proportion of education expenditure in GDP is

relatively close across most regions or countries. However, the large gap between the maximum and minimum values indicates that in some countries, there are significant deviations from the average. It is generally believed that increasing education expenditure can improve education quality and accessibility, thereby promoting higher enrollment rates.

Compulsory education duration (x2). This variable refers to the minimum mandatory education period stipulated by the government, measured by the minimum number of years of compulsory education required by national education laws and regulations, usually in years. On average, the average duration of compulsory education is nearly 10 years, reflecting the level of compulsory education protection in most regions or countries in the sample. The variance is 2.721, indicating significant differences in the number of compulsory education years across regions or countries. Some regions may implement longer compulsory education, such as 12-15 years, while some countries may not have compulsory education at all. Extending compulsory education usually increases children's schooling time and promotes higher enrollment rates, especially in primary and secondary education stages. Additionally, longer compulsory.

References

[1] Benchea, L. (2021). Analysis on the relationship between internet use and students' academic engagement. *Journal of e-Learning and higher education.*

[2] Jackson, C. K., Johnson, R., & Persico, C. (2015). The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms. *In National Bureau of Economic Research.* https: //doi.org/10.3386/w20847

[3] Olubunmi Kayode Ayanwoye. (2023). Effects of Teachers' Professional Development on Students' Academic Achievement. *Qeios.* https: //doi.org/10.32388/01sp81

[4] *Roser M. and Ortiz-Ospina E.* (2016) - "Education Spending", OurWorldinData, https: //ourworldindata.org/financing-education

[5] Stephens Jr, M., & Yang, D. Y. (2014). Compulsory education and the benefits of schooling. *American Economic Review,* 104(6), 1777-1792

[6] Løken, K. V. (2010). Family income and children's education: Using the Norwegian oil boom as a natural experiment. Labour Economics, 17(1), 118-129.

[7] Díaz Serrano, L., & Pérez Reynosa, J. H. (2018). Impact of Duration of Primary Education on Enrollment in Secondary Education: Panel Data Evidence from Developing Countries.

[8] UNESCO Institute for Statistics via UN SDG Indicators Database, UN Department of Economic and Social Affairs, https: //unstats.un.org/sdgs/dataportal/database

[9] UNESCO Institute for Statistics (UIS), *Education*, 2024, https: //uis.unesco.org/bdds

[10] *World Bank's World Development Indicators (WDI)*., https: //datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators

Appendices

Table 1: Descriptive variables

Variable name	Label
y_1 prienroll	the proportion of students in a country or region who receive education in the age range of enrollment (usually 6-12 years old)
y_2 secenroll	the proportion of students in middle and high school education.
y_3 terenroll	the proportion of students' participation at the university level and above.
x_1 spen	Public spending on education as a share of GDP (%), includes expenditures from various levels of government on school construction, teacher salaries, educational resources, and curriculum development.
x_2 eduyear	Duration of compulsory education (year), the minimum mandatory education period set by the government.
x_3 highincome	Dummy variable, =1 if it is high-income or upper-middle-income country, =0 if it is low-income or lower-middle income-country
x_{4i} teacher	Percentage of qualified teacher
x_{5i} internet	Proportion of schools with access to the internet for pedagogical purpose

Table 2: Descriptive statistics

VARIABLES	N	mean	sd	min	max
y1_prienroll	300	102.6	9.277	66.97	137.6
y2_secenroll	281	92.19	19.72	24.12	137.0
y3_terenroll	253	47.16	27.39	3.655	116.2
x1_spen	307	4.279	1.551	0.864	13.48
x2_eduyear	307	9.938	2.721	0	15
x3_highincome	307	0.720	0.450	0	1
x4i_teacher	307	93.77	13.78	1.329	100
x5i_internet	307	73.42	34.69	0	100
x4ii_teacher	307	4.529	0.281	0.845	4.615
x5ii_internet	307	4.008	1.092	0	4.615
y1i_Igprienroll	300	4.636	0.0880	4.219	4.932
y2i_Igsecenroll	281	4.506	0.259	3.224	4.927
y3i_Igterenroll	253	3.659	0.730	1.538	4.764

Table 3: The relationship between variables $x_{1\text{-}5}$

	x1 spen	x2 edu~r	x3 hig~e	x4i te~r	x5i in~t
x1 spen	1				
x2 eduyear	-0.00380	1			
0.947					
x3 highinc~e	-0.2171*	0.2528*	1		
0.000100	0				
x4i teacher	-0.1546*	-0.0422	0.1483*	1	
0.00660	0.461	0.00930			
x5i internet	-0.1599*	0.1437*	0.6413*	0.2804*	1
0.00500	0.0117	0	0		

Table 4: The relationship between variables $x_{1\text{-}5,}$ with x_4 and x_5 in logarithmic form

	x1 spen	x2 edu~r	x3 hig~e	x4ii t~r	x5ii i~t
x1 spen	1				
x2 eduyear	-0.00380	1			
0.947					
x3 highinc~e	-0.2171*	0.2528*	1		
0.000100	0				
x4ii teacher	-0.0698	-0.0609	0.0787	1	
0.223	0.288	0.169			
x5ii inter~t	-0.1496*	0.3279*	0.5930*	0.1539*	1
0.00870	0	0	0.00690		

Table 5: OLS(Ordinary Least Squares) regression results comparison with x_4 and x_5 logarithmic form & models x_4 and x_5 are not in logarithmic form

	y1_prienroll	y2_secenroll	y3_terenroll	y1_prienroll	y2_secenroll	y3_terenroll
x1_spen	-0.0498	2.1154***	3.5160***	-0.1358	2.1144***	3.4915***
	(0.3528)	(0.6609)	(1.0701)	(0.3512)	(0.6609)	(1.0531)
x2_eduyear	-0.1328	1.3673***	-1.0424*	-0.1791	1.9384***	-0.3918
	(0.2086)	(0.3286)	(0.5747)	(0.2003)	(0.3204)	(0.5456)
x3_highincome	-1.6488	13.0834***	23.0266***	-1.1185	11.3319***	19.6350***
	(1.5011)	(2.3565)	(3.7144)	(1.6031)	(2.4953)	(3.8732)
x4ii_teacher	-2.4801	4.2585	7.4113			
	(1.9258)	(3.0533)	(5.1397)			
x5ii_internet	-0.3120	7.0016***	8.2997***			
	(0.6299)	(1.0025)	(1.6344)			
x4i_teacher				-0.1006**	0.1002	0.1958
				(0.0403)	(0.0678)	(0.1231)
x5i_internet				-0.0128	0.2172***	0.2809***
				(0.0208)	(0.0326)	(0.0518)
_cons	117.7558***	12.8774	-40.6424*	116.0968***	30.5141***	-16.9110
	(9.1426)	(14.4104)	(23.8947)	(4.6361)	(7.5690)	(12.7274)
Adj R-squared	0.0056	0.4779	0.3555	0.0246	0.4798	0.3745
N	300	281	253	300	281	253
Prob > F	0.2482	0.0000	0.0000	0.0304	0.0000	0.0000

Standard errors in parentheses

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 6: OLS regression results of models with y_1, y_2, y_3 in logarithmic form

	y1i_Igprienroll	y2i_Igsecenroll	y3i_Igterenroll
x1_spen	-0.0006	0.0215**	0.0656**
	(0.0033)	(0.0086)	(0.0270)
x2_eduyear	-0.0011	0.0282***	0.0351**
	(0.0019)	(0.0042)	(0.0140)
x3_highincome	-0.0082	0.1281***	0.5061***
	(0.0153)	(0.0324)	(0.0995)
x4i_teacher	-0.0010**	0.0008	0.0041
	(0.0004)	(0.0009)	(0.0032)
x5i_internet	-0.0001	0.0031***	0.0077***
	(0.0002)	(0.0004)	(0.0013)
_cons	4.7504***	3.7381***	1.7201***
	(0.0442)	(0.0984)	(0.3268)
Adj R-squared	0.0161	0.4890	0.4190
N	300	281	253
Prob > F	0.0822	0.0000	0.0000
Prob > chi2	0.0000	0.0000	0.5687

Standard errors in parentheses

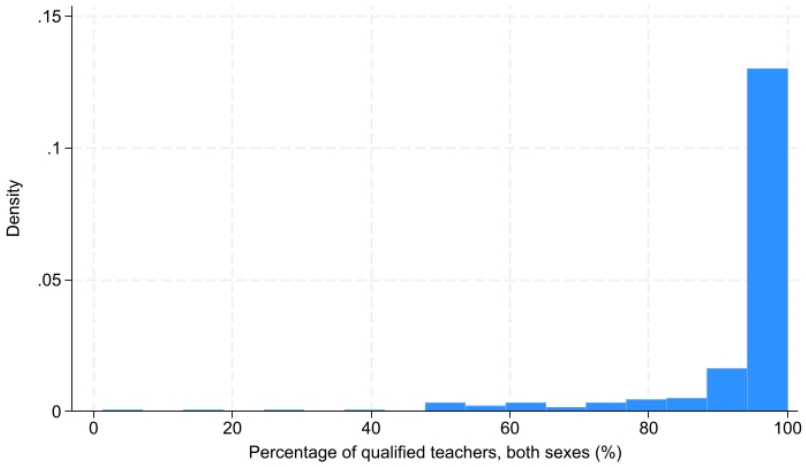
* p < 0.1, ** p < 0.05, *** p < 0.01

Table 7: Robust regression results of models with y_1, y_2, y_3 in logarithmic form

	y1i_Igprienroll	y2i_Igsecenroll	y3i_Igterenroll
x1_spen	-0.0006	0.0215**	0.0656***
	(0.0033)	(0.0089)	(0.0214)
x2_eduyear	-0.0011	0.0282***	0.0351**
	(0.0027)	(0.0052)	(0.0138)
x3_highincome	-0.0082	0.1281***	0.5061***
	(0.0125)	(0.0257)	(0.1032)
x4i_teacher	-0.0010**	0.0008	0.0041
	(0.0004)	(0.0009)	(0.0034)
x5i_internet	-0.0001	0.0031***	0.0077***
	(0.0002)	(0.0005)	(0.0013)
_cons	4.7504***	3.7381***	1.7201***
	(0.0575)	(0.1074)	(0.3348)
R-squared	0.0325	0.4981	0.4305
N	300	281	253
Prob > F	0.1788	0.0000	0.0000

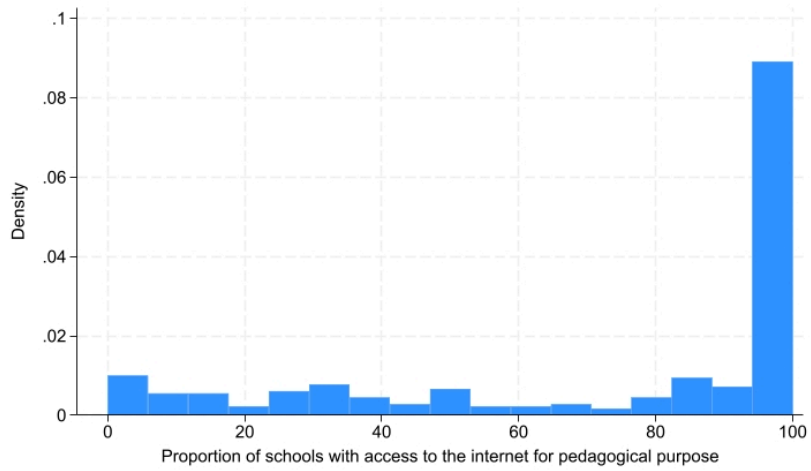
Standard errors in parentheses

* p < 0.1, ** p < 0.05, *** p < 0.01

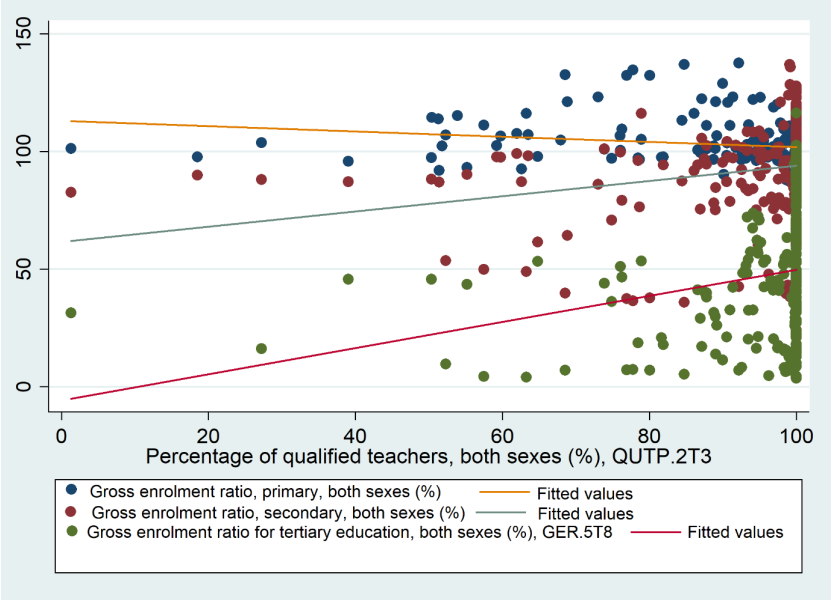


Graph 1: Histogram of percentage of qualified teachers

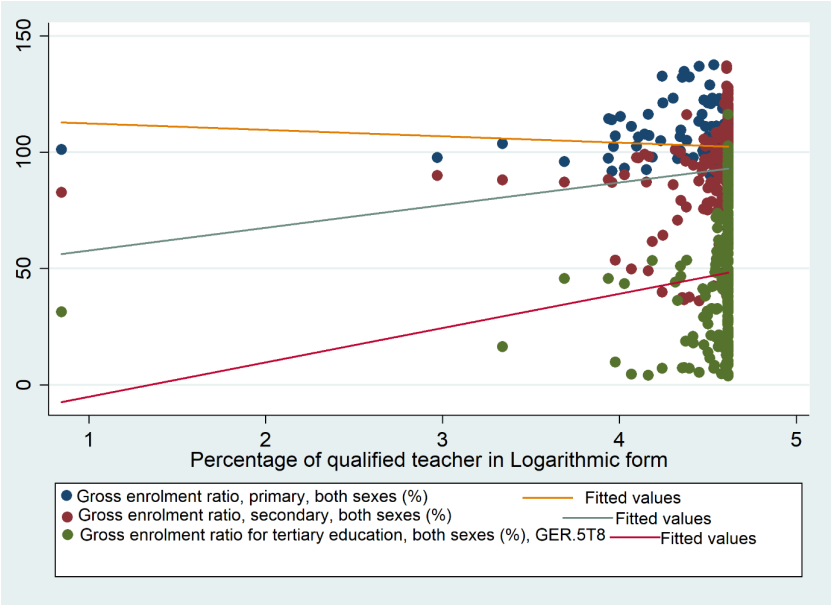
Graph 2: Histogram of Proportion of Schools With Access to the Internet



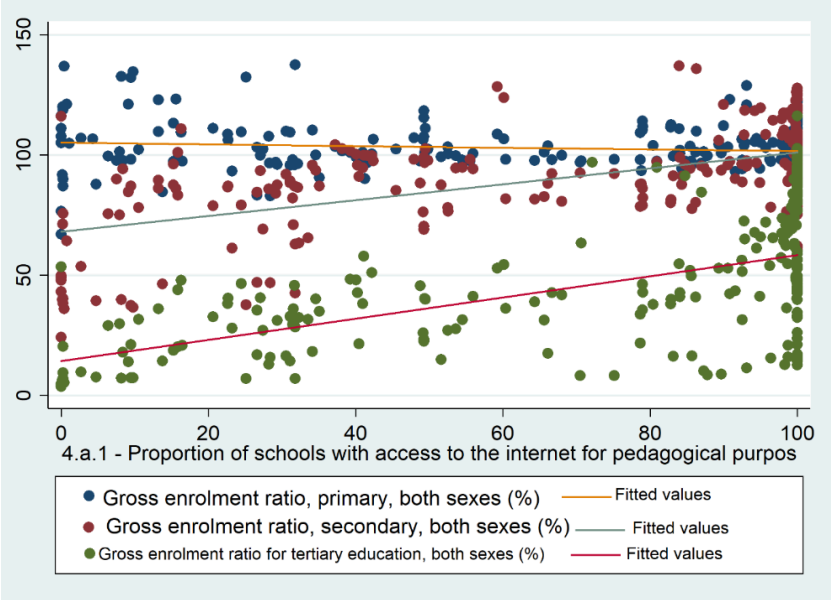
Graph 2: Histogram of proportion of schools with access to the internet



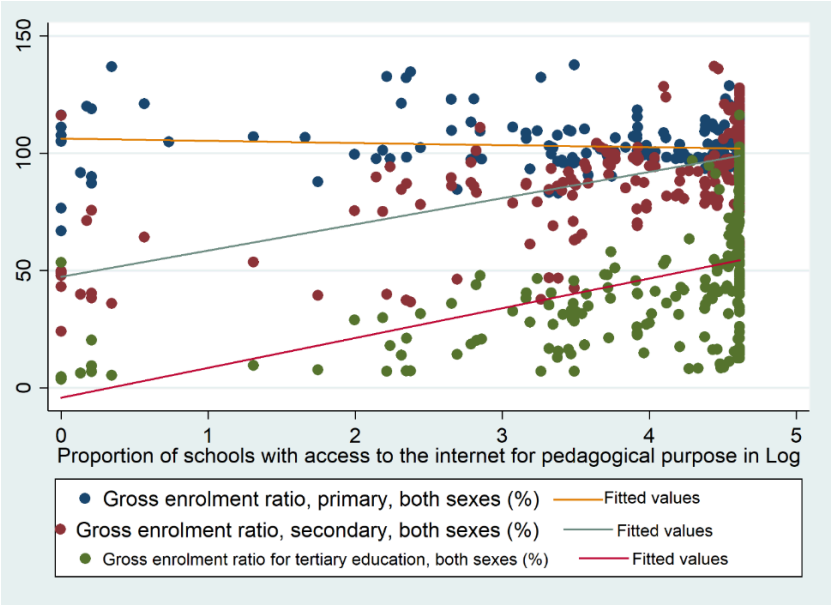
Graph 3: The relationship between gross enrollment ratio in different education levels and percentage of qualified teachers



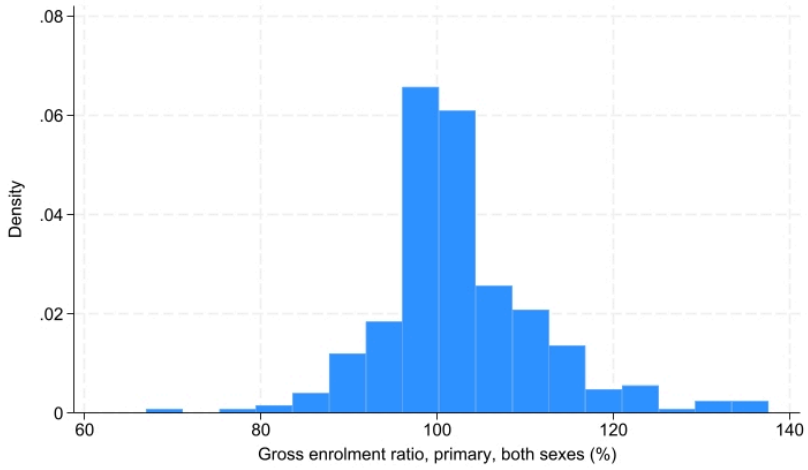
Graph 4: The relationship between gross enrollment ratio in different education levels and percentage of qualified teachers in logarithmic form



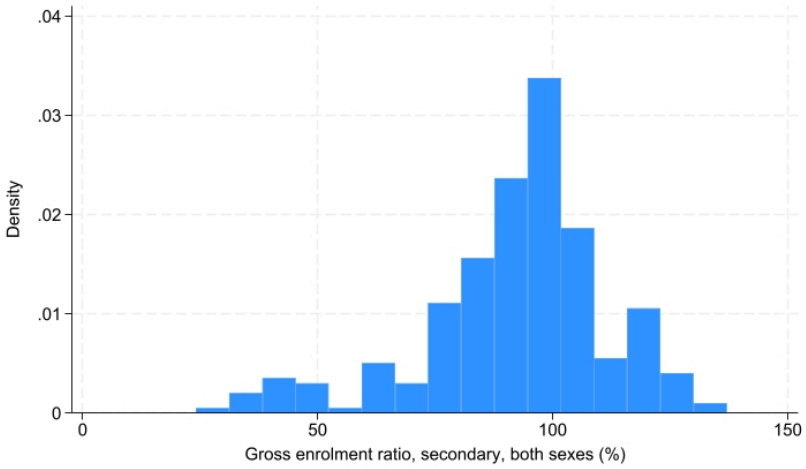
Graph 5: The relationship between gross enrollment ratio in different education levels and proportion of schools with access to the internet



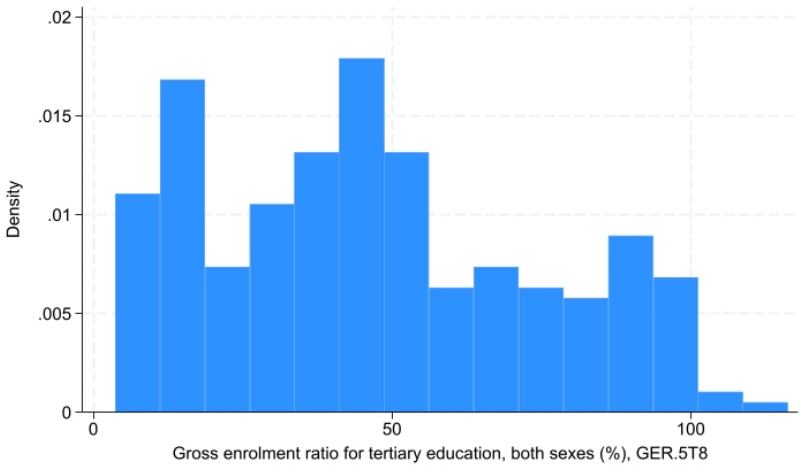
Graph 6: The relationship between gross enrollment ratio in different education levels and proportion of schools with access to the internet



Graph 7: Histogram of gross enrolment ratio for primary



Graph 8: Histogram of gross enrolment ratio for secondary



Graph 9: Histogram of gross enrolment ratio for tertiary