Spatial inequality, urbanisation and learning in Peru: evidence from the reform period of 2006-16

Sonja Fagernäs[1], Panu Pelkonen[1], Juan Manuel del Pozo Segura[1,2]

20 May 2023 (Preliminary, work in progress)

Abstract

A set of significant educational reforms swept Peru during the decade of 2010, leading to substantial improvements in learning outcomes. This study focuses on the developments in spatial inequality of primary schooling over the period, and the role of urbanisation in learning gains. Many aspects of the reforms explicitly favoured remote regions, and improvements in school inputs, such as infrastructure and teachers, were larger in remote and rural areas. Despite large general learning gains, spatial inequality in learning remained large and possibly even widened somewhat, and urban areas saw somewhat faster improvements in learning than rural areas. We analyse the urban premium in learning in public sector schools with a number of methods, and conclude that roughly half of it is due to existing differences in school resources, and large part of the rest due to differences in average school sizes and local level of development. As a parallel development to educational reforms, rapid rural-urban migration in several parts of the country accounted for a significant part of the national improvement in the learning outcomes, as pupils moved to better-resourced schools and environments. With pupil level panel data, we find that the value added in reading and mathematics is slightly larger for those who move from a rural primary to urban secondary, compared to comparison groups.

1 – University of Sussex

2 – National Foundation for Educational Research (NFER)

Corresponding author: Sonja Fagernas (s.a.e.fagernas@sussex.ac.uk)

This research has been funded by the British Academy.

1 Introduction

In the 2010s, Peru initiated a number of educational reforms, and increased funding for education accordingly. These reforms were propelled by robust economic growth, that made increased financing feasible, a notably poor performance in international rankings such as PISA, which increased the sense of urgency, and by a government that was able to reach a political consensus and implement the reforms.

While the educational system of Peru has been under constant development, the changes seen during Alan Garcia's government over 2011-16, stand out in many ways. During this period, funding for education nearly doubled, numerous targeted interventions were implemented, and objective measures of learning improved substantially. The period under study is interesting also in an international viewpoint, as an example of an effort to overhaul the education system, with tangible results.

Budget increases targeted rural areas disproportionately, given that these had been more disadvantaged to begin with. On the other hand, the more urban areas are growing more rapidly in terms of population and thus possibly also educational needs. One important question is how the period of reform has affected educational inequality between Peru's administrative regions, as well as between urban and rural areas.

The purpose of our study is two-fold. Firstly, we study the evolution of inequality in school resources, investments and learning, both in the regional and urban-rural dimension. We focus on primary education, and the public sector exclusively, given that the reform targeted the public sector. Secondly, we investigate the impact of urbanisation more specifically on learning over this time period. We rely on a range of data sets and methods. Our tests scores are primarily for second graders, but we also utilise a panel data set tracking children between grades 2 and 8.

Our analysis shows that while school resources and learning overall improved, learning outcomes across regions or between urban and rural areas did not converge. The resource gap between urban and rural areas became smaller, but urban schools continued to be better resourced, with the exception of pupil-teacher ratios. Despite higher pupil-teacher ratios, urban areas saw slightly faster improvements in learning than rural areas. The largest improvements in learning outcomes took place in regions with rapid urbanisation. Overall, learning in more rural areas benefitted from the larger increases in funding, but other factors have contributed to improved learning in urban areas, thus perpetuating the learning gaps.

We use both regression analysis and Oaxaca-Blinder decompositions to assess the size and causes of the urban learning premium. The geocoding of schools allows us to create a measure of urban density within a small 2km radius of each school, providing a much more precise assessment of the degree of urbanisation compared to the simple urban/rural division. We find that roughly half of the urban premium can be explained by school resources, such as basic infrastructure, pupil-teacher ratio and multigrade teaching. The rest appears to be largely explained by indicators relating to household socioeconomic status at the district level, private school presence, which can be indicative of wealth, but also competitive pressure and larger average school size in more urban areas. We discuss the extent to which these determinants of inequality can be influenced by policy.

From the perspective of an individual, rural-urban migration can also play a role in enhancing learning. Given that many of the rapidly urbanising regions also experienced the biggest improvements in learning, we proceed to analyse the extent to which the aggregate improvement in basic skills over 2009 to 2016 is attributable to migration of pupils from rural to urban areas. This is done via a simple decomposition analysis. The analysis shows that a sizeable share of the improvement in learning is explained by urbanisation. A similar conclusion can be drawn from a further analysis of a more selected panel data set of pupils in the second and eighth grades. Moving from a rural primary school to an urban secondary school is associated with a small increase in value-added in learning, in both reading and mathematics.

The period coincided with rapid rural-urban migration in several parts of the country, and thus pupils moving to better-resourced schools and environments, better performing peers and more educational choice. Secondly, the reform period was also characterised by robust economic growth. While the growth benefitted many kinds of areas in the country, the income gap between central and remote areas remained high in terms of income ratios, but even increased as measured by monthly incomes and expenditures. Given our focus on the public sector, we abstract here from the choice between private and public sector, which is also more relevant for urban areas.

Despite the relevance of urbanisation to economic development, there are surprisingly few studies that explicitly study how urbanisation affects learning and schooling with micro-level data. A recent study by Maarseveen (2021a) shows that children growing in urban regions in the Netherlands consistently attain higher levels of human capital compared with children in rural regions, conditional on observed cognitive ability and various family characteristics. In a working paper, Maarseveen (2021b) studies how childhood urban exposure raises primary school completion, school attendance, and literacy rates in African countries, using census data. The design is similar to Chetty and Hendren (2018), who study how neighbourhood effects affect outcomes of children who move during their childhood. While there is research at the micro level on the effects of rural-urban migration on individuals generally, this has tended to focus on adults and there is surprisingly little on the impact of children's education. We focus on the impact of urbanisation more broadly. The strength of our analysis lies in the fact that we are able to exploit a range of data sources, on schools, test scores and household data and thus explore different channels of effects.

The paper is organised as follows. Section 2 describes the Peruvian education reforms. Section 3 describes the data and descriptive statistics on regional developments in schooling. Section 4 examines urban-rural differences using a range of methods, and Section 5 focuses on the role of rural-urban migration in learning gains. Section 6 concludes.

2 Background

Description of the educational reforms

This subsection gives a compact overview of the key educational reforms that took place in Peru over approximately 2011-2016. We don't cover details of individual reforms, nor provide an evaluation of them. Instead, we will examine the package of reforms as an 'event' and its effects on regional inequality. While several of these reforms started towards the end of the period, the education budget started to grow more rapidly since 2011 and expanded by 75% between 2011-2016.

The different components of the Peruvian reforms have been studied and assessed, despite the fact that disentangling the effects of different parts of the reform can be difficult when a package of reforms is carried out over a short period. Furthermore, the reforms coincided with a substantial reduction in poverty and economic growth, which can also indirectly improve educational performance, or resources in families. It is therefore challenging to isolate the effects of individual educational reforms and disentangle them from the role of general economic improvement.

A broader context on the political economy of Peruvian educational reforms is provided by Balarin (2021), who emphasises that the weak institutions governing politics and administration has led educational reform in Peru to be characterised by discontinuity and incrementalism. Balarin concludes that while substantial reforms have taken place, significant inequalities and segregation remain, and that it is difficult to link the reforms to improvements in learning with confidence. It is clear that the school system was under targeted improvements in many ways, which will make it difficult to disentangle the role that individual programmes had on outcomes of interest. Nevertheless, several studies have evaluated the impacts of individual reforms, as explained below.

The key reforms of the 2011-16 government are documented for instance by Saavedra and Gutierrez (2020), who divide the improvements to following categories: 1) upgrades to infrastructure, 2) teacher-related reforms, 3) curricular and pedagogical reforms, 4) management of school systems. All of these were accompanied by substantial increases in financing.

Some of the reforms were explicitly aimed at improving schooling in the rural areas. For example, infrastructure spending targeted rural areas, as did the teachers' pedagogical support program *Acompanamiento Pedagogico Multigrado*. It is also likely that standardised lecture materials provided under *Soporte Pedagogico*, despite being criticised as undermining teacher autonomy, were being used more by schools with a shortage of teacher resources or less qualified teachers, both more common in remote areas.

2.1 Upgrades to infrastructure

The state of school infrastructure in Peru has been and is inadequate as some key statistics in this report will show. However, during the period we examine, there has been a steady improvement in basic school infrastructure. This was also one of the aims of the governments of the period. Funding for school infrastructure has been growing, and most of the improvements have targeted rural areas.

Specific programs targeting school infrastructure were the 'Program of Maintenance of Educational Infrastructure', which since 2012 has directed financing for schools for basic repairs and maintenance of school infrastructure. A national program to close the infrastructure gaps, the 'National Educational Infrastructure Program' (PRONIED) started in 2014. This program had financial autonomy and aimed for a standardised and quick model of monitoring and construction.

2.2 Teacher-related reforms

Key legal changes in teacher-related reforms are the 2007 *Ley de Carrera Publica Magisterial* and especially the 2012 law *Ley de Reforma Magisterial*, which opened the possibility to change how teachers are hired and promoted, based on performance evaluations. The first teacher promotion contests took place from 2014, in which teachers were able to apply for promotion and salary increase based on a written test. Such reforms, while not substantially changing the body of teachers, created incentives for self-development and made the career more rewarding financially.

Individual schools were also rewarded for good performance, based on the *Bono Escuela* program from 2014-15. A non-trivial salary bonus was given to principals and teachers of schools that performed in the top 1/3 of comparable schools in the region, based on a performance formula that accounted for learning outcomes, enrolment and retention. Bono Escuela has been evaluated by Leon (2016), who concludes that the program had positive effects on test scores in mathematics, and that it improved the attendance of teachers and school directors.

In 2014, another intervention, the A*companamiento Pedagogico Multigrado* involved visits providing teachers feedback and continuous coaching on pedagogical practices in single-teacher and multigrade primary schools - typically small schools in remote areas. Majerowicz and Montero (2022) evaluate a 2016 randomised reassignment of the program and find that the teachers trained by this program improved their pupils' test scores by over 0.2 standard deviations.

2.3 Curricular and pedagogic reforms

A new national curriculum was developed, beginning in 2010, and involving a broad consultation period over 2012-2016 with public institutions, teachers, experts and other stakeholders. The new curriculum was approved in June 2016. The new national curriculum created some political controversy due to its proposed changes in gender equality and recognition of sexual minorities. Importantly however, the curricular reform may not have affected learning outcomes much prior to the 2016, since the new curriculum was adopted only in urban areas in 2017, and later in the rest of the country.

Some changes in the curriculum did however take place earlier. Physical education and strengthened instruction of English were introduced in 2014 and 2015, respectively. Furthermore, the ministry of education developed standardised lecture materials, with additional training community activities in a program called *Soporte Pedagogico*.

Provision of pre-schooling experienced some expansion over 2011-16, especially in more rural areas. This has led to non-trivial improvements in second grade test scores in mathematics and reading, according to an evaluation by Majerowicz (2015).

2.4 Improvements in school management

The role of school principals was strengthened from 2015 onwards, when about 1/3 of existing principal positions were assigned on a meritocratic basis. Principals were also given more autonomy, for example with regards to minor maintenance purchases.

Further to this, from 2015 data collection from schools and use of data was improved under a tool, *Semaforo Escuela*, which recorded information on school functioning and attendance based on school visits and teacher interviews.

3 Data & Regional dimension of reforms

3.1 Description of the data

The analysis in the report focuses on the years 2007-2016, which coincide with the administrations of Ollenta Humala (2006-2011) and Alan Garcia (2011-2016).

With respect to school data for 2007-16, we use the Censo Escolar for 2007-2016, which is an annual school level census data set and includes information of school resources and teachers among other things. It covers all primary schools in Peru, with more than 5 pupils. We link these data to cross-sectional pupil level test score data from the Evaluación Censal de Estudiantes for the same years. This includes nationally comparable, annual test score data for second grade pupils in Reading and Mathematics[1]. In addition, we have a separate data source that includes panel data on test scores and characteristics of pupils in grades 2 and 8. Among other things, the school census data includes school level aggregates on resources, teachers, pupils and location, but does not contain information on household or parental characteristics. Data on these will be used at a more aggregated level based on the sources described below.

We use data on public finances for education from the Portal de transparencia del Ministerio de Economía y Finanzas (MEF) and compute regional expenditure on primary education for years 2005-19. Information on population and household characteristics were obtained from the Encuesta Nacional de Hogares (ENAHO) for years 2005-2019, and the national Census, Censo de Población y Vivienda, for 2007 and 2017. The former is representative at the regional (departemento) level, whereas the latter is representative at the district level. Peru has 25 regions and around 1,800 districts.

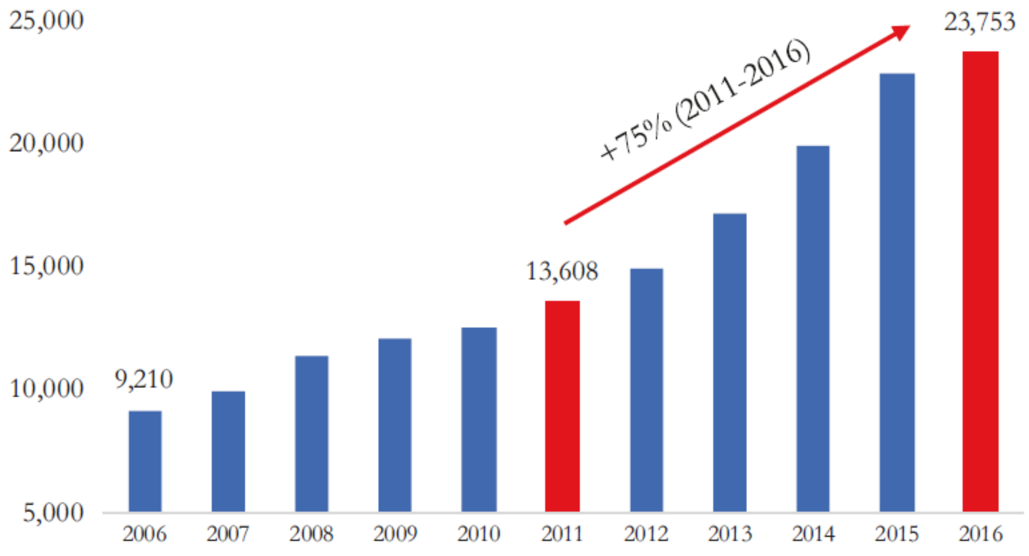
3.2 Financing

One straightforward way to assess the regional dimension of the reform period is to break down the government spending on primary education by region. Peru consists of 25 regions (or *departamentos*), at various levels of development and distance to the capital Lima.

Figure 1 shows the significant increase in the total education budget, in particular in the latter part of the period, 2011-2016. Figure 2 reveals a similar picture for spending on primary education specifically. Figure 3 shows that the increases in spending were larger in more rural areas: there is a negative association between the rise in spending between 2012-2016 and the share of pupils in urban schools in 2011 at the region level. Figure 4 reveals a similar picture with respect to learning deficiencies: there is a negative association between reading score of grade 2 pupils in 2011 and consequent increase in spending per pupil between 2012-2016. These figures indicate that spending increases were directed more towards areas lagging behind in educational achievement.

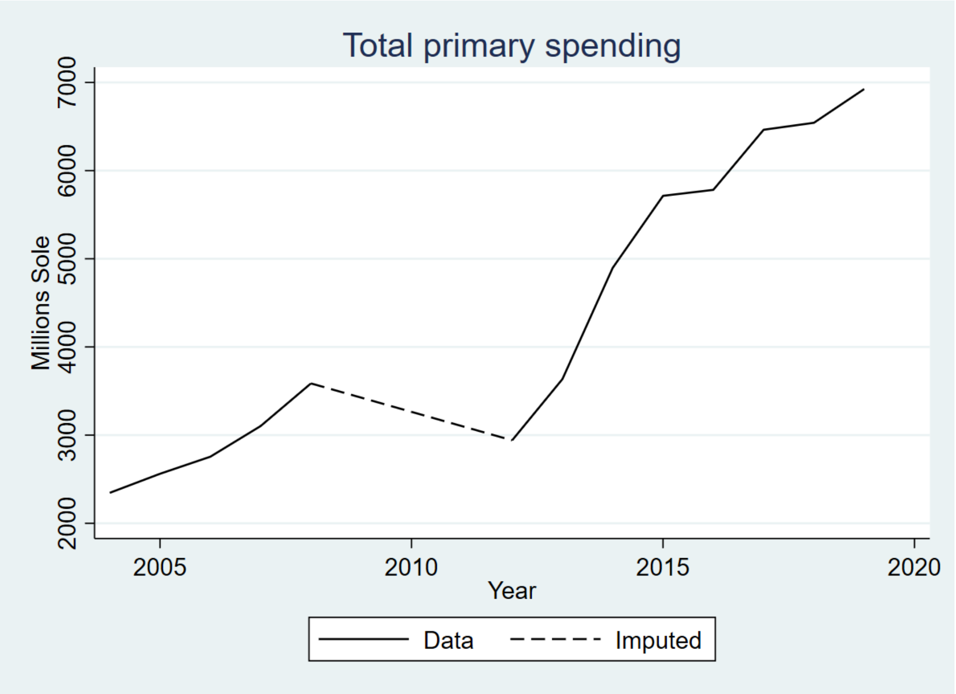
[1] Bilingual schools test their pupils in grade 4, and we do not include them in our sample as the test results are not directly comparable.

Figure 1 Total budget allocated to education in 2006-2016, millions of Soles.



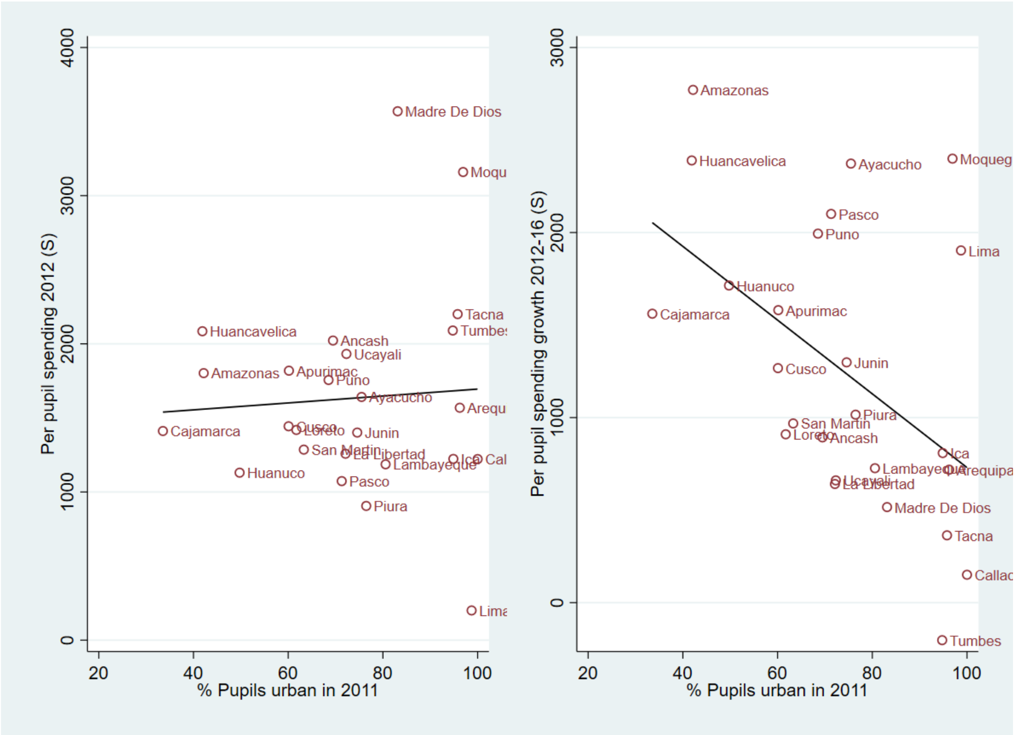
Source: Saavedra and (Gutierrez 2020)

Figure 2 Total budget allocated to Primary education 2004-2019, Millions of Soles



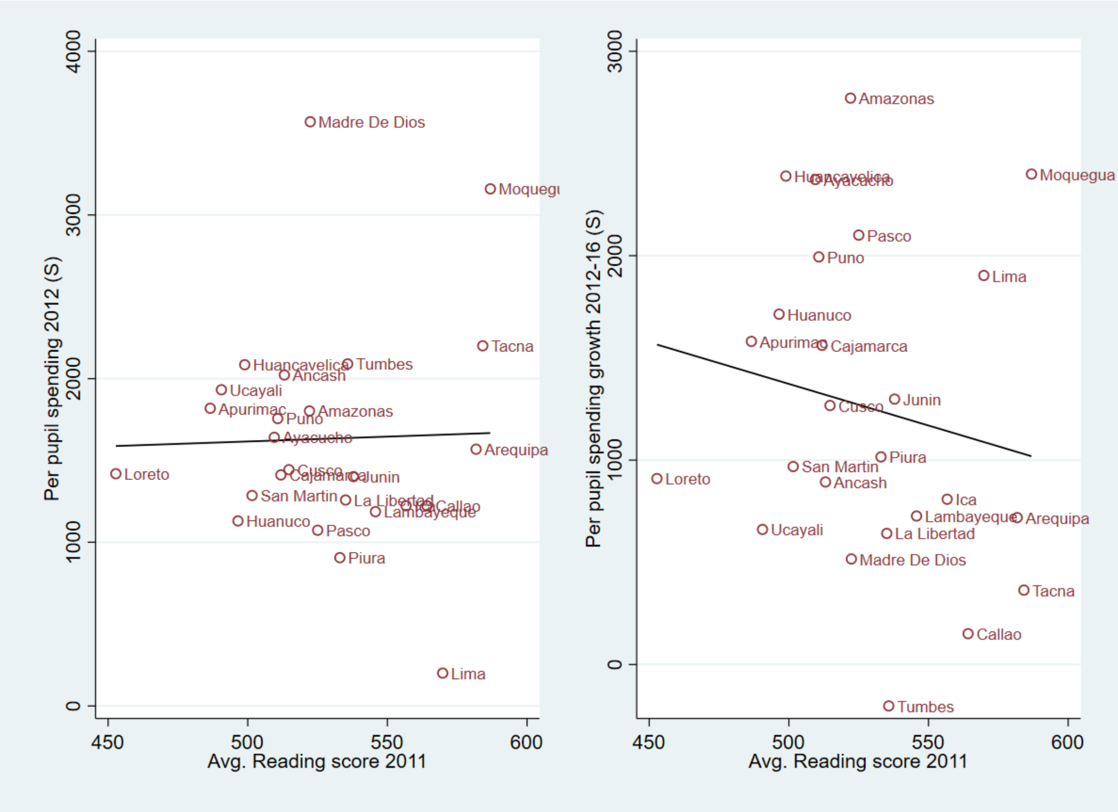
Notes: Source: Own calculations using MEF Budget data. Years 2009-11 are interpolated due to missing data or changes in definitions.

Figure 3 Funding improvement by urbanisation level of region over 2012-2016



Notes: The funding figures in y-axes are in current Soles and are based on annual planned regional budget for primary education. It has been scaled to 'per pupil' by dividing by annual number of public sector pupils. The x-axis is the proportion of school census pupils in region that are in urban schools in 2011 (includes private sector pupils). The black line is an unweighted linear fit.

Figure 4 Funding improvement by literacy level of region over 2012-2016



Notes: The funding figures in y-axes are in current Soles and are based on annual planned regional budget for primary education. It has been scaled to 'per pupil' by dividing by annual number of public sector pupils. The x-axis is the average reading score of second grade pupils 2011 (includes private sector pupils). The black line is an unweighted linear fit.

3.3 Patterns of regional convergence and divergence

Table 1 below presents some key statistics for public primary schools for the 25 regions of Peru. The statistics are for the years 2007 and 2016, which represent the starting and end points for our analysis. The statistics relate to the average scores for Reading and Mathematics in grade 2, the average pupil-teacher ratio (PTR), the share of pupils in single-teacher schools, in urban schools, in private schools and in bi-lingual schools[2], where the language of instruction is not Spanish. We also include an indicator for the availability of basic five infrastructure items in schools (Water, Sewage, Electricity, Toilet, Internet), ranging between 0 to 5. The numbers reported are the pupil-weighted means in the whole region.

[2] We exclude the bi-lingual schools from the statistical and econometric analysis later on as the pupils were tested at a later stage (in Grade 4). These school tend to be concentrated in more rural areas.

Table 1 Summary statistics of pupils, teachers, urbanisation and infrastructure for 25 regions in 2007 and 2016.

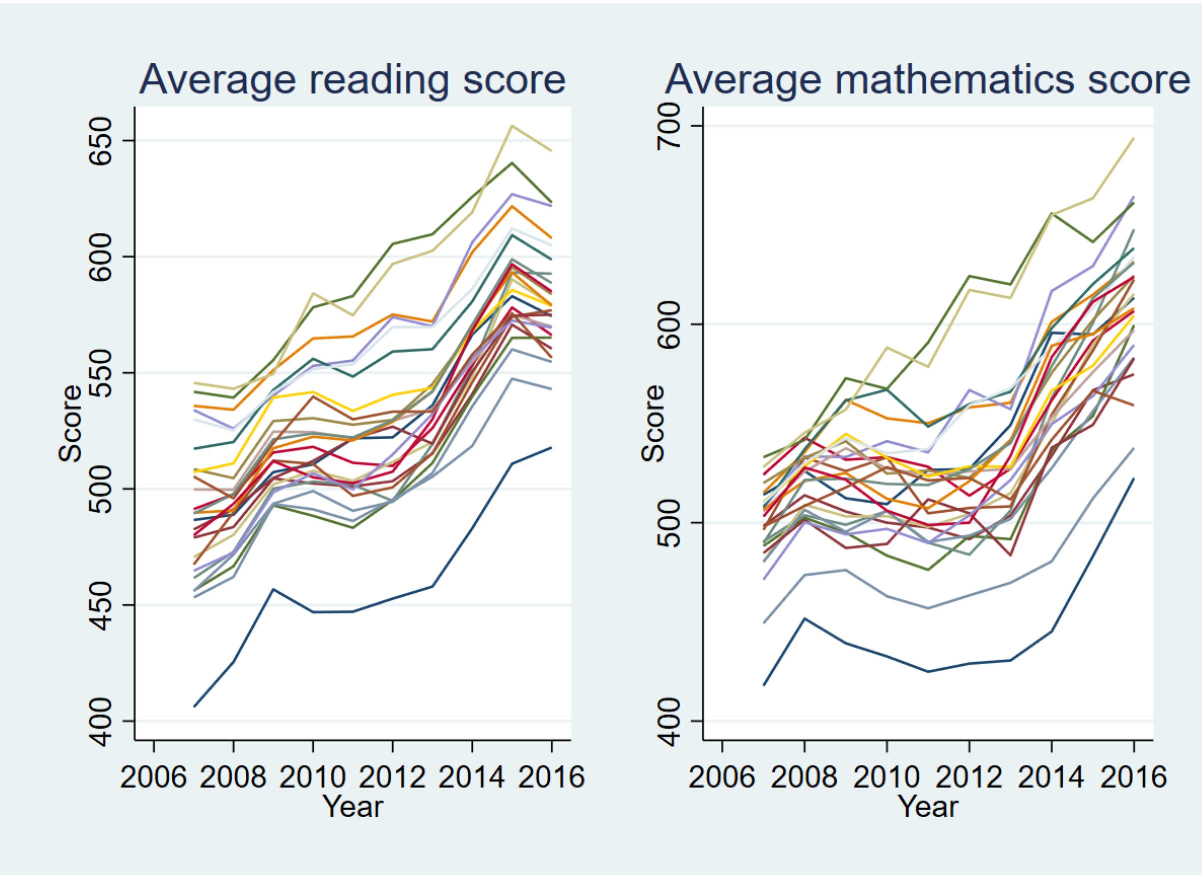
Region	Year	Reading	Maths	PTR	Infra 5	Single T	Urban	Private	Bilingual
Amazonas	2007	487	514	25.0	2.2	0.58	0.38	0.02	0.23
Amazonas	2016	575	613	20.0	4.1	0.35	0.63	0.02	0.06
Ancash	2007	488	503	26.1	3.3	0.25	0.61	0.11	0.06
Ancash	2016	566	573	21.5	4.4	0.19	0.75	0.16	0.10
Apurimac	2007	458	489	25.4	2.5	0.32	0.48	0.02	0.04
Apurimac	2016	567	597	20.6	4.1	0.22	0.70	0.06	0.17
Arequipa	2007	555	529	23.0	3.9	0.08	0.91	0.38	0
Arequipa	2016	608	615	19.9	4.6	0.06	0.97	0.41	0
Ayacucho	2007	467	493	25.9	3.0	0.35	0.60	0.06	0.10
Ayacucho	2016	596	642	19.9	4.3	0.16	0.79	0.11	0.15
Cajamarca	2007	494	525	25.6	2.0	0.53	0.30	0.05	0
Cajamarca	2016	568	604	21.9	4.1	0.39	0.58	0.07	0
Callao	2007	542	514	27.6	4.1	0.01	1.00	0.32	0
Callao	2016	611	633	23.9	4.8	0.03	1.00	0.40	0
Cusco	2007	479	486	28.0	3.0	0.35	0.54	0.11	0.05
Cusco	2016	583	611	21.6	4.3	0.16	0.75	0.16	0.14
Huancavelica	2007	470	498	25.6	2.4	0.60	0.30	0.02	0.31
Huancavelica	2016	577	620	17.8	4.1	0.30	0.62	0.04	0.20
Huanuco	2007	459	482	29.2	2.3	0.49	0.37	0.06	0
Huanuco	2016	559	582	22.2	4.1	0.29	0.66	0.10	0.03
Ica	2007	526	514	24.4	3.3	0.07	0.90	0.21	0
Ica	2016	597	621	23.1	4.8	0.07	0.97	0.28	0
Junin	2007	516	525	27.5	3.1	0.25	0.62	0.12	0
Junin	2016	588	619	21.6	4.5	0.21	0.83	0.20	0.08
La Libertad	2007	511	511	27.4	3.1	0.26	0.69	0.18	0
La Libertad	2016	575	590	24.1	4.4	0.19	0.81	0.23	0
Lambayeque	2007	518	514	28.2	3.3	0.16	0.78	0.20	0.02
Lambayeque	2016	582	594	24.1	4.7	0.13	0.89	0.25	0.03
Lima	2007	543	520	26.3	4.2	0.03	0.94	0.37	0
Lima	2016	600	604	22.6	4.8	0.04	0.99	0.47	0
Loreto	2007	411	422	26.5	2.2	0.36	0.55	0.05	0.03
Loreto	2016	521	524	25.0	3.6	0.31	0.69	0.06	0.13
Madre De Dios	2007	486	486	29.0	3.2	0.12	0.85	0.06	0
Madre De Dios	2016	574	576	28.9	4.1	0.11	0.87	0.12	0.03
Moquegua	2007	547	537	21.5	4.2	0.05	0.81	0.16	0
Moquegua	2016	625	656	19.3	4.7	0.03	0.98	0.20	0
Pasco	2007	492	508	23.5	3.4	0.30	0.60	0.05	0
Pasco	2016	582	609	19.3	4.3	0.21	0.79	0.09	0.07
Piura	2007	498	495	29.3	3.2	0.27	0.64	0.13	0
Piura	2016	591	624	26.7	4.4	0.23	0.83	0.19	0
Puno	2007	485	506	25.0	2.9	0.29	0.62	0.09	0.02
Puno	2016	588	619	20.4	4.5	0.12	0.82	0.16	0.06
San Martin	2007	466	472	25.7	2.2	0.36	0.56	0.03	0.01

San Martin	2016	570	589	23.4	4.2	0.26	0.75	0.03	0
Tacna	2007	553	536	25.5	4.2	0.02	0.93	0.23	0
Tacna	2016	646	689	22.7	4.8	0.02	0.98	0.21	0
Tumbes	2007	509	499	23.0	3.9	0.02	0.89	0.11	0
Tumbes	2016	562	560	22.3	4.7	0.05	0.96	0.15	0
Ucayali	2007	456	452	27.7	2.4	0.32	0.61	0.03	0.13
Ucayali	2016	545	538	28.2	4.3	0.24	0.78	0.06	0.15

Notes: Learning outcomes are second grade results and include only Spanish-language schools. PTR = Pupil-teacher ratio, Infra5 = Pupil-weighted summation index [0-5] consisting of Water, Sewage, Electricity, Toilet and Internet, SingleT = Single teacher school, Urban = Share of pupils in urban schools, Private = Share of pupils in private schools, Bilingual = Share of pupils in bilingual schools.

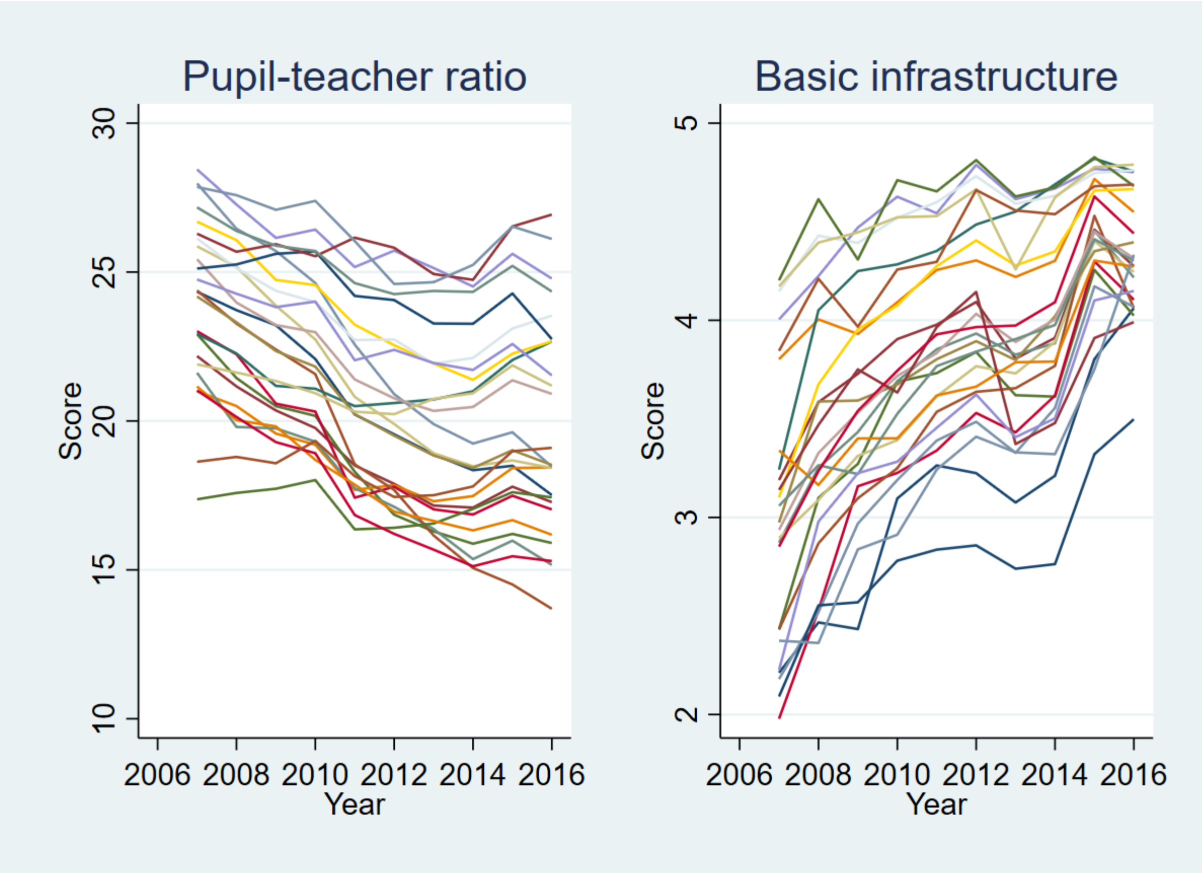
The figures show significant increases in test scores, infrastructure availability and urbanisation in most of the regions. The improved test performance across regions is also evident from Figure 5. Learning outcomes in Peru have improved across all regions over 2007-16, roughly at the same pace. This has left the regional inequalities largely unbridged, even if the overall development has been commendable. These statistics refer to outcomes of the school system at early stages of primary schooling (grade 2).

Figure 5 Average learning outcomes across 25 regions in Reading and Mathematics, grade 2 pupils, 2007-2016. Public Spanish-language schools.



In Figure 6, we summarise the regional development in two key dimensions of school inputs: teachers, and basic school infrastructure. Teacher input is summarised as the average number of pupils per teacher in the region, and the basic infrastructure as a simple index from 0 to 5 which counts how large share of pupils are in a school with Water, Sewage, Electricity, Toilet and Internet. A notable feature of Figure 6 is that teacher inputs appear to have diverged across regions, implying that the attention pupils get from teachers is becoming more unequal over time. The infrastructure index on the other hand, suggests convergence – the regions that were behind the most at the beginning of the time frame, have caught up with the leading regions to some extent.

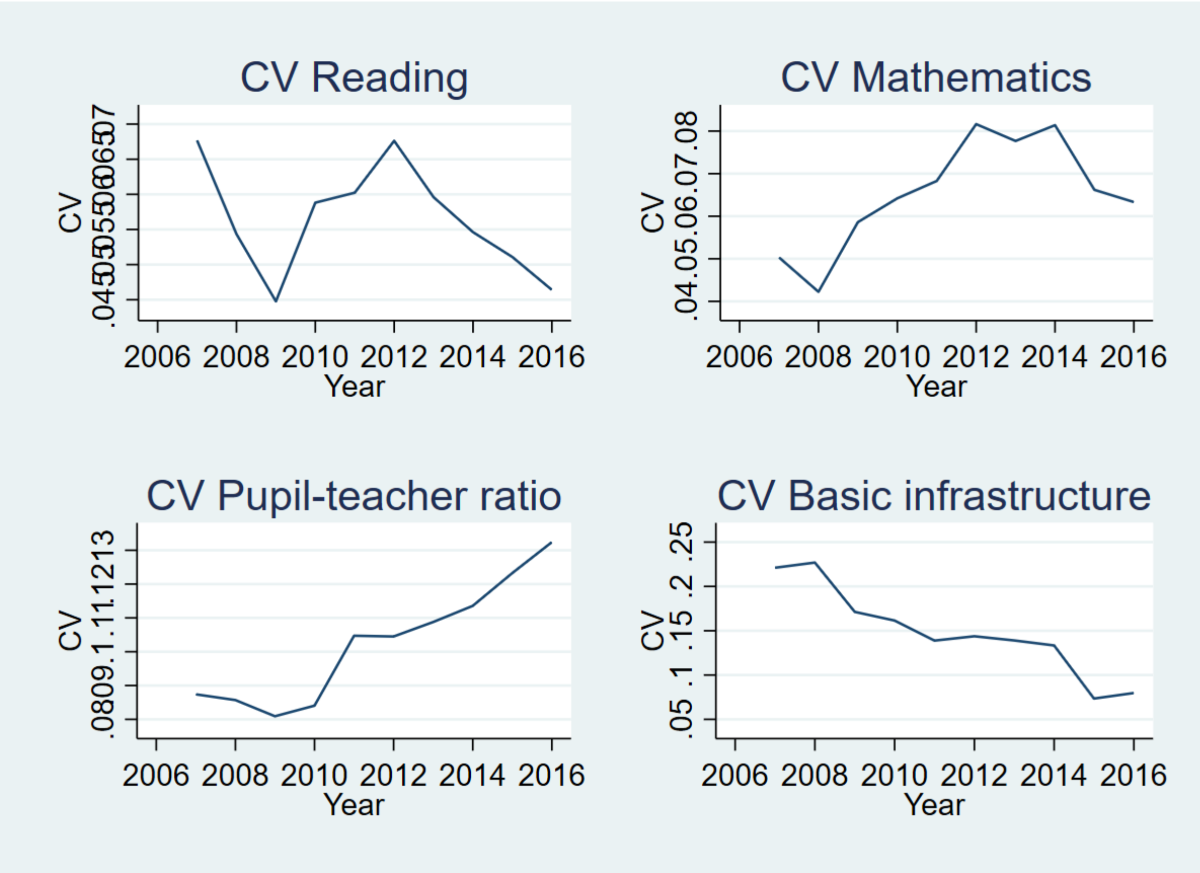
Figure 6 Developments in key school inputs in public primary schools over 2007-2016.



Notes: Basic infrastructure is a sum of five items (Water, Sewage, Electricity, Toilet and Internet), and has been averaged across schools in the region using pupil number weights. Pupil teacher-ratio has been calculated for the whole region (public sector only).

Developments of inequality over time can also be summarised by computing a coefficient of variation (mean / standard deviation) over the years for variables of interest (Figure 7). In terms of inputs (bottom images in Figure 7), the coefficients of variation show that the teacher inputs have become regionally more unequal over time, whereas physical infrastructure has become more equal. With respect to learning, regional inequality has declined somewhat since 2012, especially in Reading, but as will be seen below, this does not imply a reduction in achievement gaps for instance between rural and urban areas.

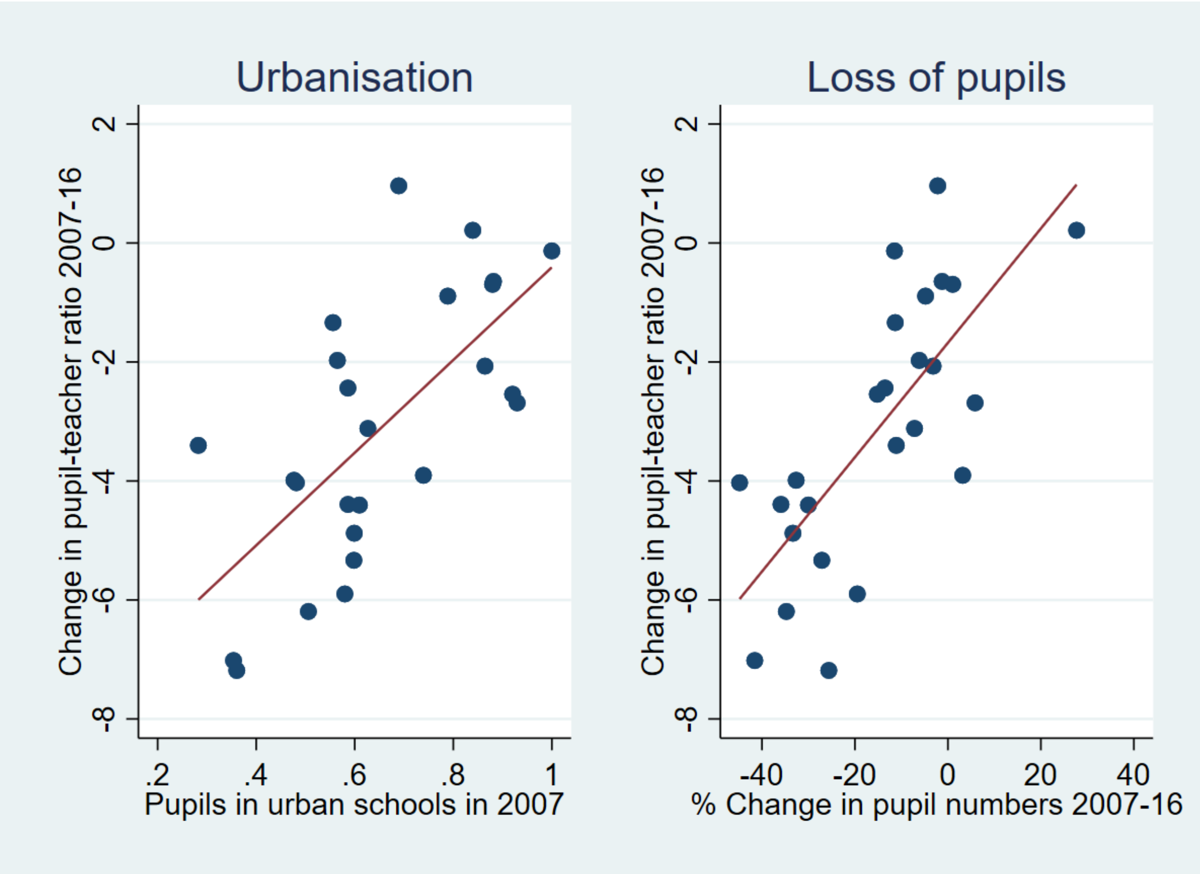
Figure 7 Coefficients of variation over time for learning outcomes and key inputs, 2007-2016.



It is worth investigating in further detail, why pupil-teacher ratios across regions are diverging. A significant part of the answer lies in urbanisation, and in how the numbers of teachers are adjusted in schools, as pupil numbers change. Figure 8 shows two scatter plots. The first one shows the share of pupils in urban schools in 2007, and the subsequent change in the pupil-teacher ratio over 2007-16, across all 25 regions. This demonstrates that in most urban regions, the pupil-teacher ratio has remained roughly constant, whereas in more rural regions, pupils per teacher have declined.

The second scatter plot in Figure 8 shows the total change in the region's primary pupil population, plotted against change in the pupil-teacher ratio. This reveals a strong positive correlation: the regions that lose pupils, see a reduction in pupil-teacher ratio, implying that the number of teachers doesn't decline at the same rate the number of pupils. Regions that are relatively stagnant with respect to pupil numbers, also see reductions in their pupil-teacher ratio, but to a lesser degree. This is due to increases in teacher numbers. The regions that experience fastest growth in pupil numbers, do not experience improvements in pupil-teacher ratios, since teacher recruitment is only just keeping up with pupil numbers.

Figure 8 What explains changes in pupil-teacher ratios across regions over 2007-16?



Notes: Scatter plots of 25 regions. Changes are computed over 2007-2016.

These developments can be interpreted as an equalising force that supports schools in declining rural regions. However, since this has not been accompanied by catch-up in learning outcomes, a more pessimistic interpretation is possible as well: it is difficult to terminate contracts of teachers or transfer them to other schools, potentially leading to structural imbalances in the allocation of teachers.

3.4 The role of the private sector?

The share of pupils in private primary schools has increased slightly over the period in urban areas, until 2014 (Figure 9). A similar pattern can be seen for regions (Figure 10).

While our focus is on the public sector, given its significant presence, it is worth exploring how the presence of the private sector contributes to inequality.

To study the association between private schooling and learning over the time period, in Table 2, we use panel data for 25 regions for 2007-2016 and explain the average grade 2 learning outcomes in Spanish language schools as a function of the share of pupils in different school types. We control for year effects to control for generic improvement in learning, and for region fixed effects to account for persistent differences in environment and demography across the regions.

Figure 9 Share of private school pupils in urban Peru

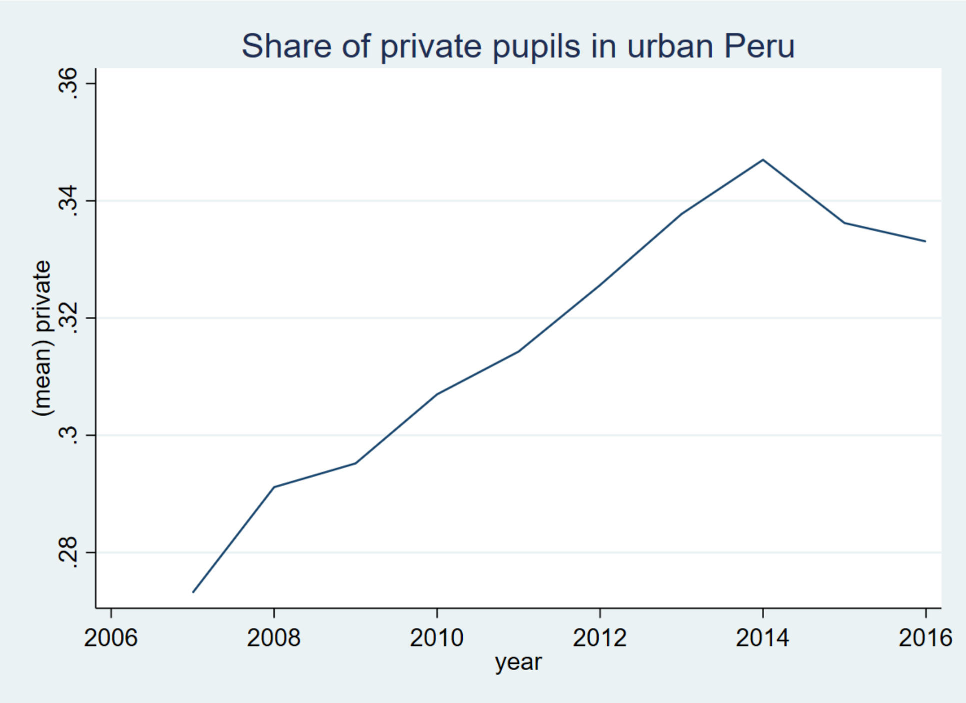
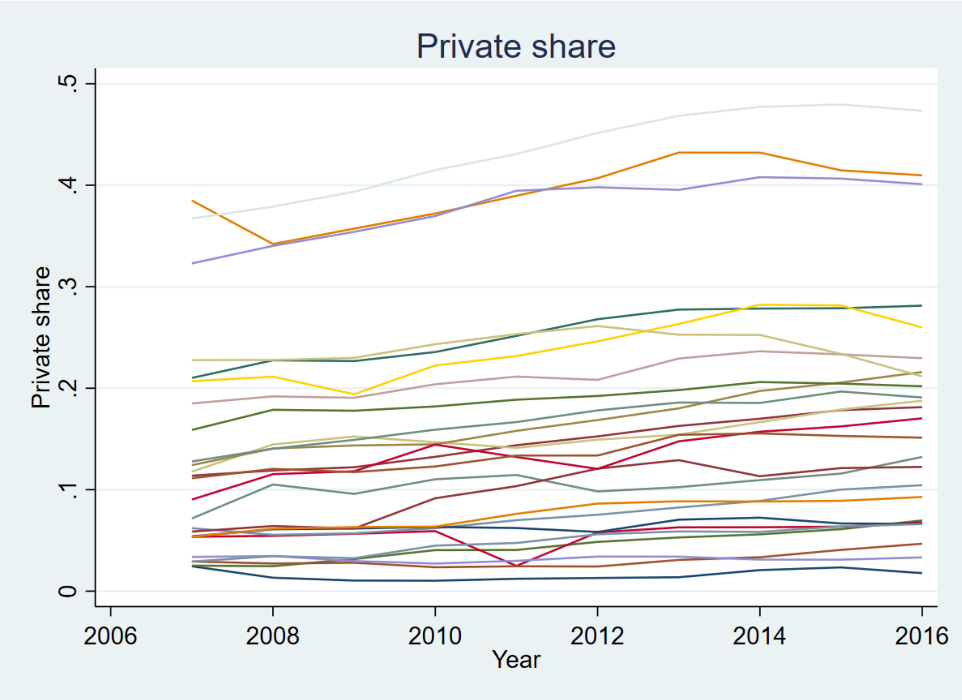


Figure 10 Share of Spanish-medium primary pupils in private schools, by 25 regions over 2007-2016.



Notes: The numbers exclude pupils in bilingual schools.

Overall, Table 2 suggests that regions where private schooling has grown over 2006-2017, have been left behind in terms of reading results. The coefficient -115, which is statistically significant, suggest that a 10-percentage point increase in privatisation would result in a 11.5 point reduction in reading score, which is about 0.1 standard deviations. For mathematics, this is less clear, since the estimated coefficient, while negative, is much smaller and not statistically significant.

Table 2 Private schooling and average grade 2 learning outcomes in 25 Regions

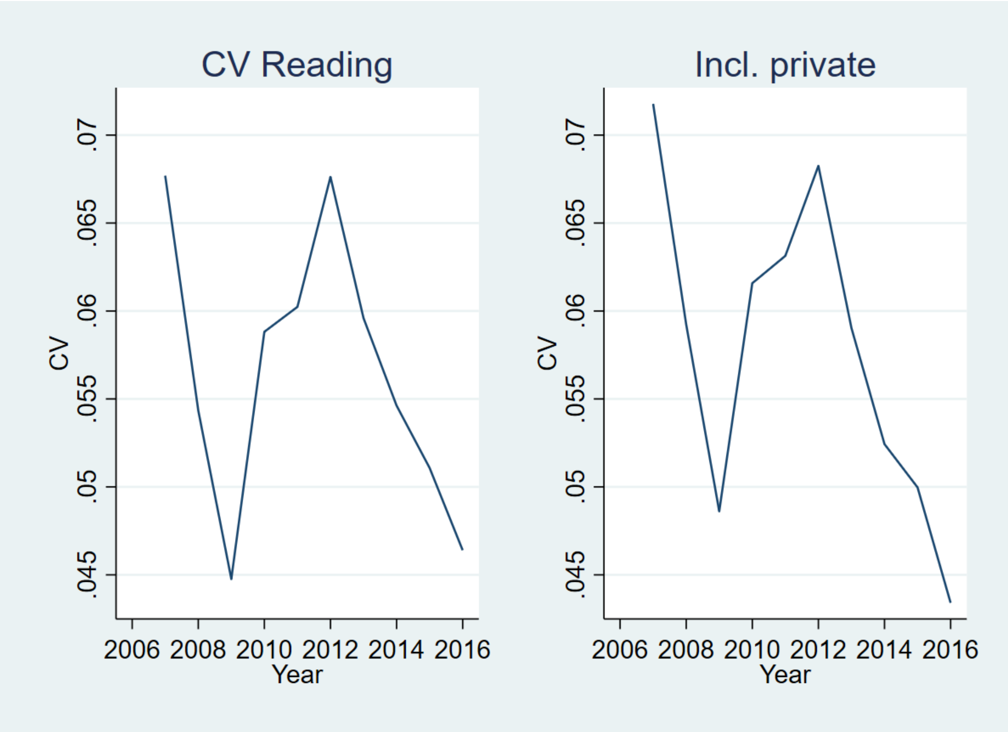
	[1]		[2]	
	Reading		Mathematics	
	coef	se	coef	se
Share in Private school	-115.108**	[41.442]	-48.097	[67.838]
Share in Bilingual school	23.171	[22.025]	-25.083	[36.053]
Year 2008	6.018**	[2.269]	20.444**	[3.715]
Year 2009	26.103**	[2.308]	24.485**	[3.777]
Year 2010	34.320**	[2.407]	23.350**	[3.940]
Year 2011	32.979**	[2.451]	18.583**	[4.013]
Year 2012	41.314**	[2.567]	24.515**	[4.202]
Year 2013	49.506**	[2.749]	31.459**	[4.500]
Year 2014	74.145**	[2.854]	63.722**	[4.671]
Year 2015	99.324**	[2.890]	84.683**	[4.730]
Year 2016	89.900**	[2.855]	105.190**	[4.674]
Observations	250		250	
R-squared	0.970		0.931	

Notes: +,*,** signify significance at 10%, 5% and 1% level. Data is region averages 2007-2016. Both models include region fixed effects. Means are computed from Spanish-speaking public and private schools. Bilingual school share is controlled, but they are not included in the dependent variable as they are not tested in grade 2.

A somewhat counterintuitive consequence of privatisation could be, that since the pace of privatisation is faster in more developed and urban regions, the process of privatisation may end up reducing the regional inequality in learning outcomes. We examine this in Figure 11, where we have computed the coefficient of variation in reading using two different samples. In the left side image, we use only public Spanish language schools, and in the right panel, we pool both public and private Spanish language schools.

When private schools are included (the right side of figure 11), we see that there is a slight tendency towards regional equalisation of results. The final datapoint in 2016 has shot down somewhat lower than in the left side panel. This is because the prevalence of (poor quality) private schooling is suppressing learning outcomes in the regions that on average have had nationally good learning outcomes.

Figure 11 Coefficient of variation in reading results across regions, with and without private schools.



Before concluding that privatisation reduces regional inequalities, it should be remembered that privatisation is also an indicator of a larger set of choices, which has been exercised by the parents. This choice has created some benefits for parents that would not be accounted for in this analysis (such as shorter distances to school).

3.5 Bilingual schools

The bilingual schools in Peru, that give instruction in an indigenous language, have been concentrated in rural areas, and the share of primary pupils in such schools has increased from about 2.4% in 2007 to about 3.7% in 2016. Being typically located in economic periphery, such schools, and how they fare, can be an important component in spatial educational inequality.

Figure 12 shows how the primary age pupil population is divided into different school types: Spanish medium public schools, private schools and bilingual public schools. The diversity has increased, as there has been a decline in the share of pupils in Spanish medium public schools. The share of pupils in bilingual schools has been growing slowly, but the growth is clearly slower than that of the private sector.

Compared to Spanish-medium public schools, 92% of the bilingual schools are in rural areas (66% for Spanish-medium), and smaller with an average of 70 pupils per school, which is about half of the typical Spanish-medium public schools (Table 3).

Figure 12 Shares of pupils in different school types

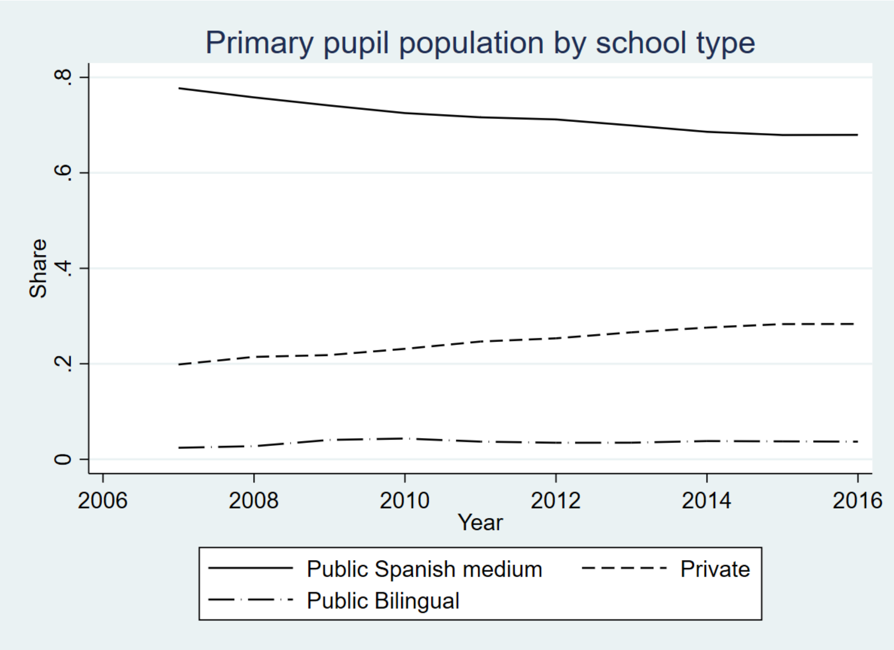
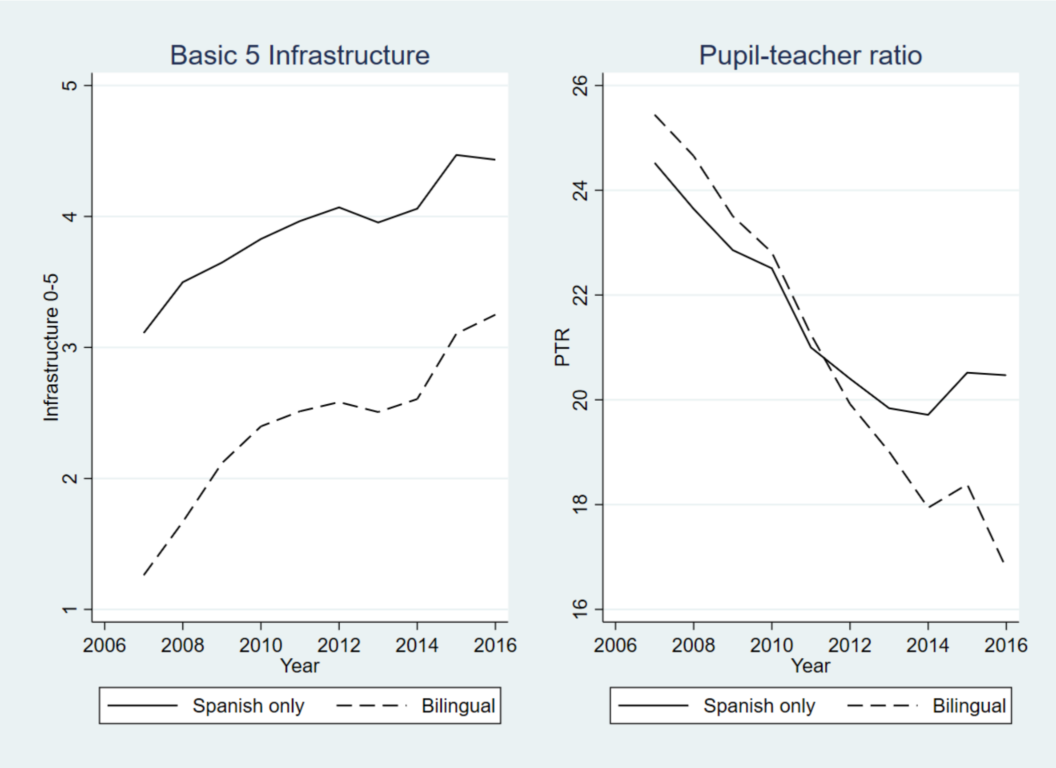


Table 3 Summary statistics for bilingual and Spanish medium public schools, 2007-2016

	Spanish language schools				Bilingual schools			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Basic 5 infrastructure	3.04	1.43	0	5	2.30	1.44	0	5
Pupil-teacher ratio	20.4	7.8	1	70	22.2	9.5	2	70
Enrolment	143.5	203.0	1	2238	70.3	51.4	2	997
Urban school	0.34	0.47	0	1	0.08	0.28	0	1
	Obs = 164239				Obs = 16535			

Figure 13 plots the two key school inputs over time in public schools by the medium of instruction: key infrastructure and the pupil-teacher ratio. This shows that the infrastructure in bilingual schools has improved somewhat more rapidly than in Spanish medium school, but by 2016 the bilingual schools are still clearly behind. Part, but not all of this difference is due to smaller school size of the bilingual schools. In the right hand side image in Figure 13, we can see that the pupil-teacher ratios in bilingual schools have caught up with the rest, and are 2016 lower than in Spanish medium schools. This development mirrors the developments between the urban and rural Spanish medium public schools examined earlier; the efforts directed at more rural parts of the school system have delivered visible gains in terms of school resources.

Figure 13 Infrastructure and PTR in bilingual and Spanish medium public schools



As mentioned earlier, the difficulty with bilingual schools is that the learning outcomes from these schools are not directly comparable with the ones from Spanish medium schools. Pupils in bilingual schools are tested in grade 4 instead of grade 2, and furthermore, are not tested in Mathematics. The pupils are tested in Spanish, and one of four indigenous languages (Asháninka, Aymara, Quechua or Shipibo).

4 Inequality in the urban-rural dimension

In this Section, we explore the development in the urban-rural disparities in schooling over the time period in question and analyse the possible factors behind the higher learning outcomes in public sector schools in urban areas.

4.1 The urban learning premium

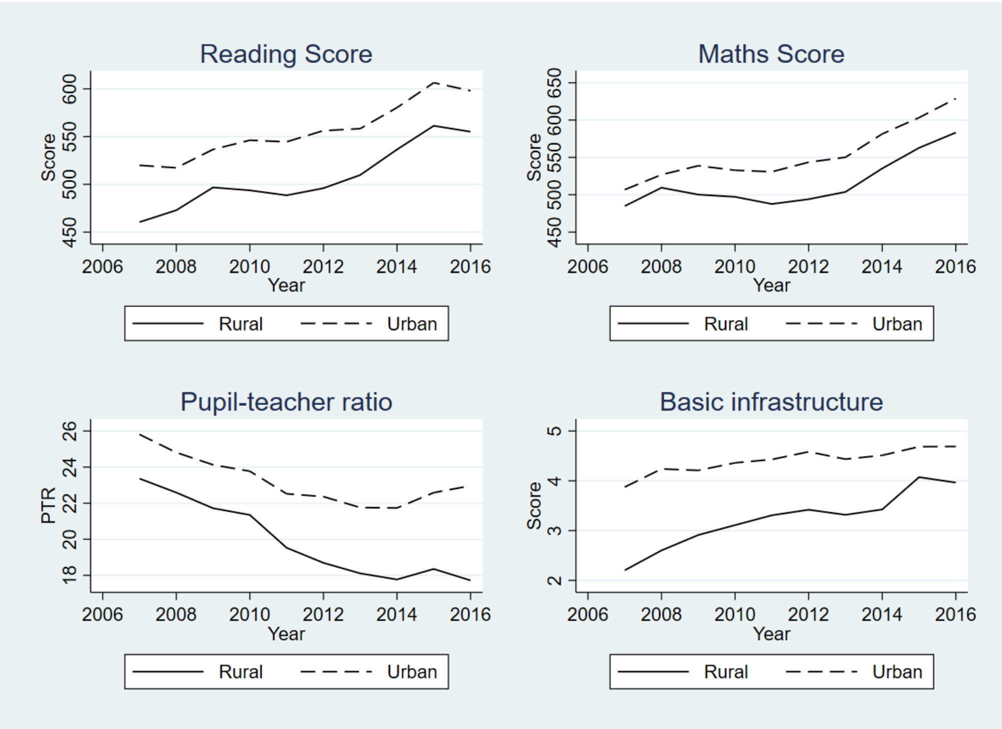
The graphs until now have only considered the 25 regions of Peru as the main spatial dimension of interest. Naturally, these regions contain a diversity of conditions and experiences, with both urban and rural and wealthy and poor areas. In this subsection, we move to a more disaggregate level with the district as the unit of analysis. A district is a sub-regional unit and there were around 1,800 districts in Peru in 2016. We focus specifically on the urban/rural distinction, as well as using a continuum of educational backwardness.

Table 4 School resources and outcomes by urban-rural status in 2007 and 2016

	2007		2016	
	Rural	**Urban**	**Rural**	**Urban**
Share of pupils in bilingual school	0.073	0.005	0.200	0.013
Reading Score	451	515	526	592
Maths Score	479	506	546	624
Pupil-teacher ratio	23.2	25.5	16.0	21.8
Share in single teacher school	0.684	0.029	0.747	0.037
Basic infrastructure (0-5)	1.71	3.84	3.40	4.57

Table 4 and Figure 14 indicate that test scores have improved since 2007 in both urban and rural areas, but rural areas have not caught up with urban areas, the gap appears to have slightly widened with respect to mathematics. Pupil teacher ratios have on the other hand declined significantly more in rural areas. Infrastructure has improved in both, again more in rural areas, but rural areas continued to lag behind urban areas.

Figure 14 Results and key school resources over 2007-2016 by urban/rural districts



Notes: Districts are definer as 'urban' is their rate of urbanisation was over 80% in Census 2007 (391 districts), and 'rural', if below 80% (1373 districts).

The earlier graphical analysis showed that school financing received a substantial boost especially after 2011, which coincides with a new government (Alan Garcia). This was also the period in which rural areas started seeing substantial increases in resources per pupil. We use an event study approach to analyse whether the period of budget increases over the government of Alan Garcia (2012-16) had a different impact depending on the degree of urbanisation at the start in 2011.

We estimate a set of the following type of models:

$$(1)\ \text{Outcome}_i = \beta \text{Garcia}_{it} \times \text{Urbanisation}_d + \lambda_i + \theta_t + \varepsilon_{i.}$$

In this specification (1), we control for school fixed effects (λ_i) and year effects (θ_t), and estimate whether the Garcia period affected schools differently depending on their district's urbanisation status prior to the term using an interaction term. If parameter β is estimated to be positive, the improvements would be larger in more urban regions, and vice versa, if negative.

The results are shown in Table 5. Firstly, as indicated by the graphs earlier, the education budget increases were smaller in areas that were more urban. More rural districts gained more in terms of resources, with larger falls in pupil-teacher ratios and more improvements in resources (Table 5, columns 1-3). Nevertheless, reading and especially mathematics scores increased more in urban regions (columns 4-5). This suggests that other factors than resources and budgets have also contributed to the improved learning in these regions, thus perpetuating the learning gaps. It is therefore worth investigating the urban learning gains further. In the Section below we analyse the development in household economic conditions between the urban and rural areas.

Table 5 The effect reforms by initial rate of urbanisation

	(1) Budget per pupil	(2) Pupil-teacher ratio	(3) Basic 5 Infrastructure	(4) Reading score	(5) Maths score
Government 2011-2016	-442.018**	3.355**	-0.545**	5.588**	26.445**
× Urbanisation	[51.671]	[0.267]	[0.046]	[1.616]	[2.690]
Year effects	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	115,436	163,686	152,715	164,676	164,643
R-squared	0.868	0.663	0.738	0.629	0.508

Notes: **: p <.01, *: p <.05, +: p <.10. Standard errors clustered by district. Urbanisation of district is from 2007 census.

4.2 Developments in household economic conditions

School resources and teachers are of course not the only determinants of learning outcomes. The economic conditions of households and parental inputs are also likely to be important factors, although more difficult to assess.

We begin with give an overview of the developments in economic conditions of households across regions and discuss their potential effect on the spatial inequality of educational outcomes. The key data sources will be censuses of 2007 and 2017, as well as the annual household survey, ENAHO, for years 2007-2016. The former is representative at the district level but the latter only at the regional level. The ENAHO is an annual, cross-sectional household survey.

Considering the basic level of housing, we collect four items from the Censuses of 2007 and 2017: The share of households that have water supply, electricity, a toilet, and sufficient amount of space per person, to the extent that the household is not considered to be overcrowded. In the vast rural areas of Peru, these necessities are far from obvious, while being potentially quite important for the learning environment of primary aged pupils. Table 6 summarises the district level means for these variables for the two censuses, as well as the sum of these four indicators, which potentially varies between 0-4.

Table 6 Summary statistics for district-level factors for 2007 and 2017

	Census 2007		Census 2017	
	Obs.	Mean	Obs.	Mean
Home Quality index (0-4)	1,831	2.136	1,874	2.916
Adequate water supply	1,831	0.458	1,874	0.750
Electricity supply	1,831	0.551	1,874	0.774
Household not overcrowded	1,831	0.854	1,874	0.915
Toilet	1,831	0.273	1,874	0.477

Source: Census 2007 and 2017

With respect to these indicators, we see a robust development. Water supply has increased by nearly 30 percentage points from low base rate of 46%. Presence of electricity and toilet have likewise increased by about 20 percentage points. The share of households not overcrowded has increased from 85% to 92%. Overall, these numbers convey a picture of rapid reduction of poverty.

Figure 15 plots the development of the sum of these basic 4 factors for the 25 regions of the country. The trajectories across all regions are positive, with slightly steeper improvements for the regions that were lagging behind in 2006. This suggests that when it comes to the very basics of housing quality, there has been some convergence, or equalisation of conditions across the regions.

Looking at averages across regions will of course hide the variability within regions, especially in the urban-rural dimension. To address this, Figure 16 plots the level of the 'household basics' indicator for the 2007 and 2017 censuses against the degree of urbanisation at the district level in 2007. This shows that while urban districts have an advantage with respect to household quality both in 2007 and 2017, there has been substantial catch-up in more rural districts.

The basic household services such as electricity and toilets cover only one crude dimension of household well-being and resources. In in Figure 17, we plot the average hourly pay across the 25 regions over 2007-2016, based on the ENAHO data. These data suggests that the story of regional convergence may not be as clear as with other indicators.

Figure 15 Basic household infrastructure across 25 regions

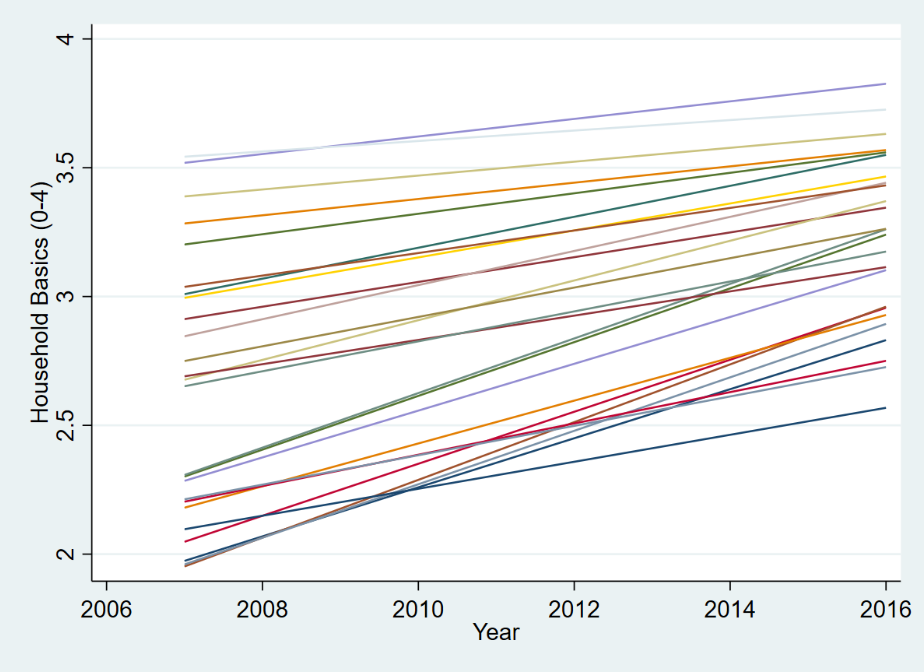
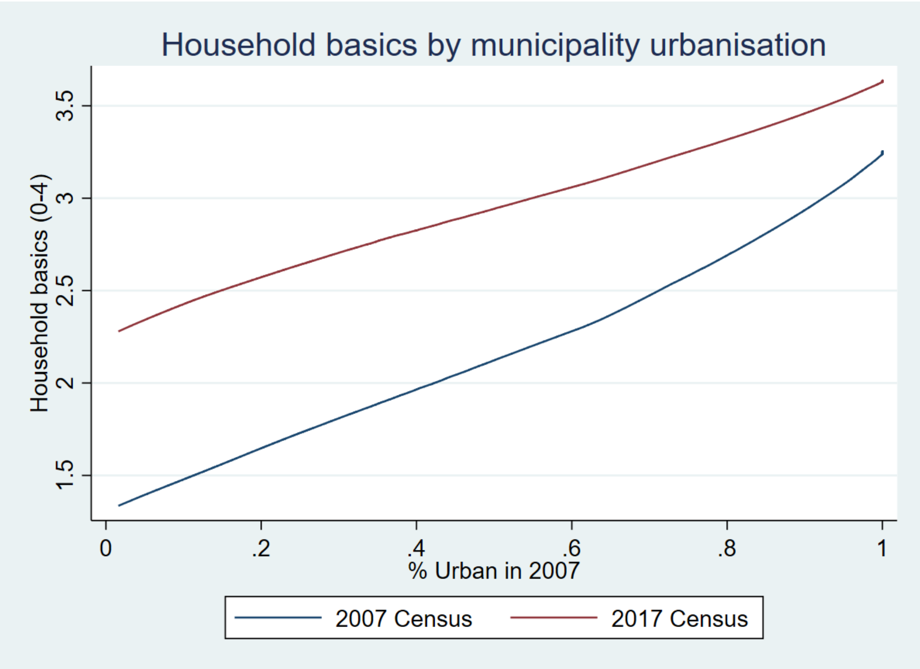
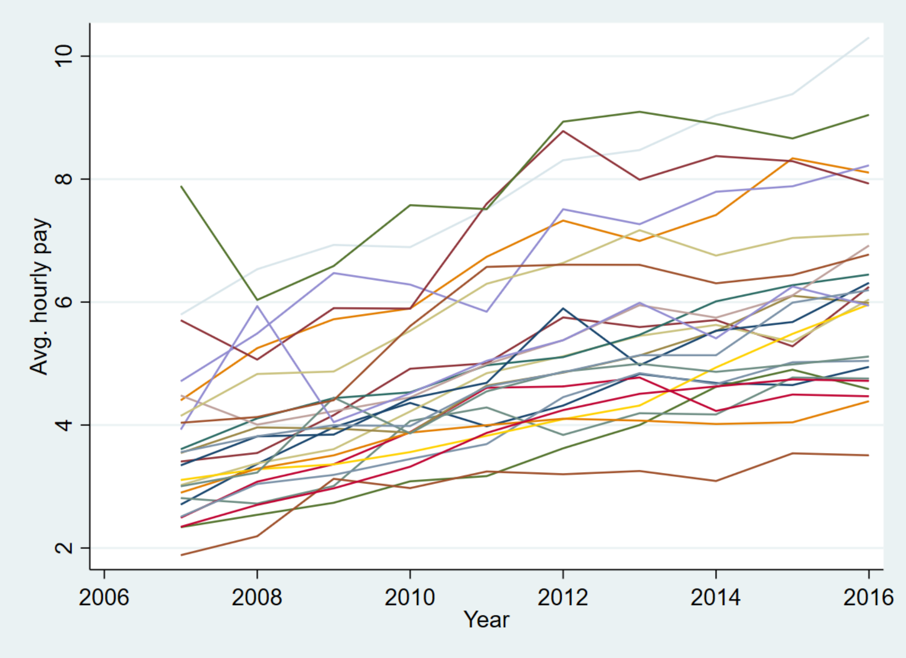


Figure 16 Catch-up in basic household infrastructure quality by initial level of municipality urbanisation.



Notes: The y-axis, 'Household basics' is a municipality sum of 4 household-level indicator variables: Adequate water supply, Electricity, Toilet, and Non-overcrowding. Source, Censuses 2007 and 2016.

Figure 17 Average hourly pay across regions, ENAHO 2007-2016.

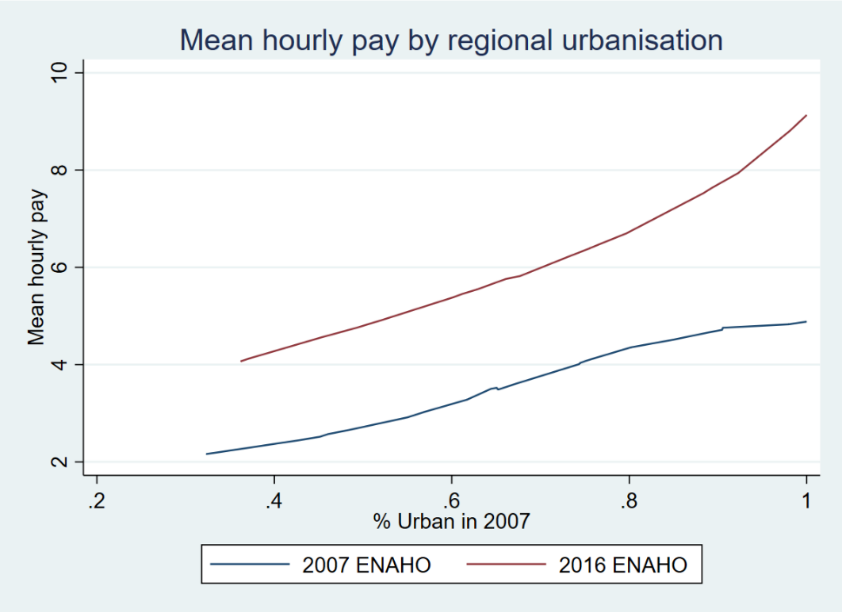


Notes: The y-axis, 'Average hourly pay' is from ENAHO household survey, averaged across regions each year in current Soles (not adjusted for inflation). By 2016, the lowest pay is in Huancavelica, and highest in Lima.

Visually, from Figure 17 it is clear that the range of average hourly pay has been expanding. For example, in Lima the mean pay grew from 5.80 Soles per hour to 10.30, while in Huancavelica, it grew from 1.88 to 3.51. These figures suggest that the pay gap between the regions may have increased. However, when looking at the ratio of incomes, in 2007, Lima's hourly pay was 3.09 times higher, but in 2016, 2.93 times higher. More generally, the growth of average hourly pay has been roughly similar across regions.

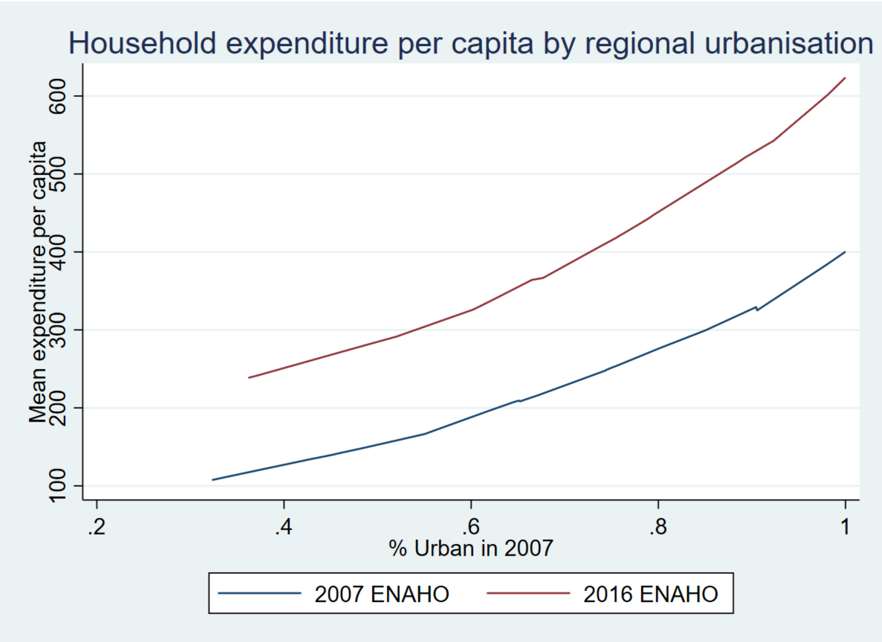
When we divide hourly pay by urban and rural location, it becomes clear that urban areas have seen more rapid growth in pay and household expenditure than rural areas (Figures 18 and 19). Finally, the income gap between the larges towns and others has grown larger since about 2012 (Figure 20).

Figure 18 Catch-up in average hourly pay by initial level of regional urbanisation.



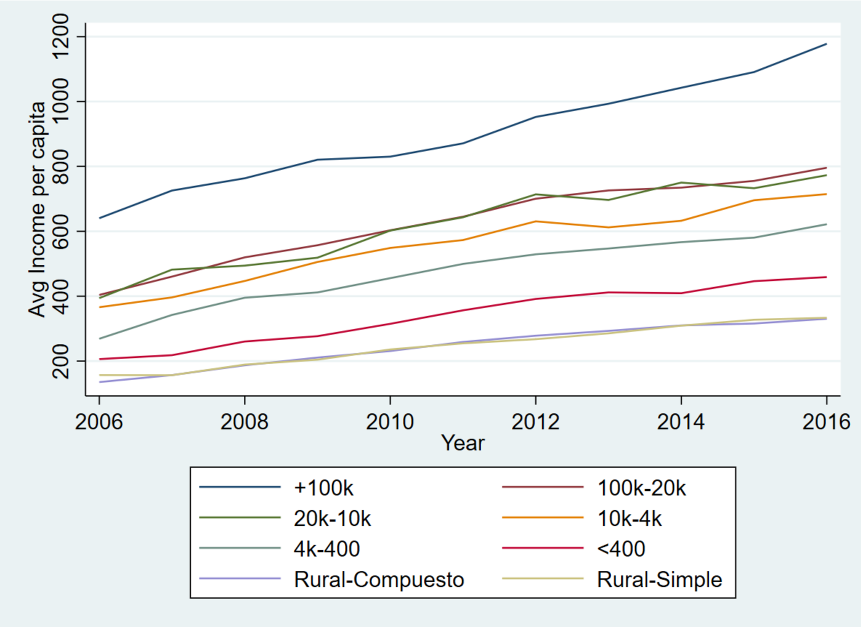
Notes: On x-axis, 25 regions are ordered by the initial rate of urbanisation in 2007. The y-axis is mean hourly pay in regions in 2006 and 2016 (Running smoothing). Data sources: ENAHO 2007 and 2016 for wages, Census 2007 for rate of urbanistion.

Figure 19 Catch-up in expenditure per capita by initial level of regional urbanisation.



Notes: On x-axis, 25 regions are ordered by the initial rate of urbanisation in 2007. The y-axis is mean household expenditure per capita in regions in 2006 and 2016 (Running smoothing). Data sources: ENAHO 2007 and 2016 for expenditure, Census 2007 for rate of urbanistion.

Figure 20 Average Income per capita by size of town/locality, 2006-2016



Notes: Source: ENAHO 2006-2016.

4.3 Evidence from population density in micro localities

In this Section, we estimate a set of models to the study the presence and determinants of the urban learning premium. We use a two-year panel data set of public schools for years 2007 and 2016 and estimate a set of regression models with and without school fixed effects with the reading and mathematics scores of second grade pupils as the outcomes of interest. The summary statistics for the variables used in the regressions are shown in Table 7.

To measure the degree of urbanisation, we rely on the population density of the local area. Given that the school census includes location information for all schools in the latest censuses, we compute the number of primary school pupils within a 2 km radius of the school, and take the natural logarithm of this number. This number varies quite dramatically from a handful of pupils to tens of thousands, and can be considered a good proxy of the density of the residential settlement in the school's immediate surroundings.

Table 8 shows the results for the reading scores and Table 9 those for the mathematics scores. Column 1 in Table 8 establishes a simple correlation based on an OLS model; more dense locations have better learning outcomes. The correlation is strong; density and the year dummy for 2016 alone explain 40% of the variation in the reading scores. On average, moving from the 10^{th} to the 90^{th} percentile in terms of density, increases reading scores by 12.54*(9.58-3.61) = 74, which is nearly a whole standard deviation (82 in the sample).

Table 7 Summary statistics for public primary schools, 2007 and 2016

	2007					2016				
	Obs.	Mean	S.D.	Min	Max	Obs.	Mean	S.D.	Min	Max
Reading score	17,987	457.9	81.1	62	740	13,782	546.6	61.6	230.5	818.6
Mathematics score	17,961	482.6	89.3	77	848	13,782	569.8	93.0	236.7	893.5
Ln Pupils within 2km	10,465	5.868	2.115	2.079	10.51	8,246	6.224	2.293	1.792	10.481
% Private within 2km	10,465	0.104	0.189	0	1	8,246	0.173	0.244	0	1
Pupil-teacher ratio	18,087	23.8	8.1	1.7	70.0	13,677	18.3	7.2	2.3	70.0
Multigrade	18,099	0.675	0.468	0	1	13,782	0.544	0.498	0	1
Single teacher	18,099	0.167	0.373	0	1	13,782	0.091	0.287	0	1
Basic 5 Infra	18,099	2.011	1.624	0	5	10,142	3.922	1.014	0	5
Ln Enrolment	18,099	4.413	1.009	0.693	7.713	13,782	4.447	1.064	1.386	7.592
% Female	18,099	0.491	0.184	0	1	13,782	0.491	0.169	0.000	1.000
Years of schooling	18,099	5.734	1.781	1.914	12.500	13,782	6.603	1.789	3.439	13.280
Home quality	18,099	2.077	0.823	0.425	3.994	13,782	2.909	0.634	0.507	3.997

Notes: School data from School Censuses. ‘Years of schooling’ and ‘Home quality’ are district-specific means computed from Censuses 2007 and 2017.

Table 8 Learning premium from local density – Reading

	[1]		[2]		[3]		[4]	
	Reading		Reading		Reading		Reading	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln Pupils in 2km radius	12.54**	[59.22]	5.11**	[16.78]	0.87*	[2.29]	-5.38*	[-2.53]
% Private in 2km radius					16.21**	[5.38]	-13.35	[-1.49]
Pupil-teacher ratio			-0.93**	[-10.41]	-0.62**	[-6.81]	-0.41*	[-2.45]
Multigrade			-7.39**	[-4.18]	-5.08**	[-2.90]	-4.00	[-1.16]
Single teacher			-3.09	[-1.43]	-5.14*	[-2.40]	11.93+	[1.90]
Basic 5 Infra			5.49**	[13.28]	4.32**	[10.35]	-0.97	[-1.57]
Ln Enrolment			14.61**	[13.59]	13.02**	[12.17]	14.84**	[5.09]
% Female			8.53**	[3.03]	8.85**	[3.18]	4.87	[1.01]
District years of schooling					4.00**	[7.82]	4.06	[1.55]
District home quality					6.06**	[5.19]	6.61**	[2.90]
Year 2016	84.11**	[89.53]	72.87**	[60.85]	68.02**	[53.44]	73.93**	[33.11]
Constant	396.82**	[285.42]	381.70**	[81.10]	368.18**	[73.65]	409.21**	[16.40]
School fixed effects	N		N		N		Y	
Observations	18,663		16,887		16,887		16,887	
R-squared	0.40		0.46		0.47		0.91	

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. Panel data with two years, 2007 and 2016.

In column 2, we add basic observable variables on school resources and school size into the regression model, which reduces the density premium by more than a half, from 12.54 to 5.11. This suggest that more than half of the urban premium is due to simple, observable school quality indicators. Reading scores are higher in larger schools, captured by enrolment numbers and schools with better basic resources. They are lower in school with multi-grade teaching, and higher pupil-teacher ratios. If we would be able to control for unobservable school quality measures, such as the differences in the quality of teaching, the drop in the density effect would most likely be even larger.

It is also the case that urban areas have richer and more educated households. At the regional level, as described in the previous section, the ENAHO data suggested that wage growth had been faster in more urban areas. The Census, which is representative at the district level does not include data on wages or incomes, but to capture education and wealth, we use data on the years of adult schooling in the district, and an index for housing quality, ranging between zero and four, accounting for electricity, water, a toilet, and overcrowding. The numbers for 2016 are taken from 2017 census. Further, we calculate the share of primary pupils that are in private schooling, within 2km radius of the public school, using the school census data. For most public schools, this share is zero, but in urban areas it quickly rises to very large shares. This variable can be an additional proxy for income, but can potentially also capture competition and selection effects created on the public schools by the presence of private schools.

Column 3 shows the results of models that include the above mentioned variables. Firstly, these controls cut the density premium further down to 0.87, or to about 7% of the original. This effect is no longer impressive, as it would only imply that moving from the 10th density percentile to the 90th percentile increases reading scores by 0.87*(9.58-3.61) = 5 points. Again, should we have more precise local information on the levels of income, wealth and education, this effect would probably be even smaller. Secondly, the coefficient on local private schooling is highly significant and large: if the share of pupils in private schools grows from 0 to 50%, reading scores increase by 8 points in public schools, or about 0.10 standard deviations. This can reflect the unobserved wealth of households in the area, but also capture the effect of competition, choice and selection. Note that if private schools capture the best students in the area, one might even expect a negative coefficient on the share of private pupils.

In the final column, we incorporate school fixed effects. This model explains changes in learning as a result of changes in pupil population density over time in the vicinity of the schools. Due to rapid urbanisation in Peru, some areas have seen quite substantial change in population density. Another force creating change in the density indicator is the changing size of pupil cohorts over time. Interpreting the resulting coefficient may be difficult because rapid change in a local area may create its own negative or positive effects on learning. The headline estimate in fact turns negative, suggesting that places growing in density start doing worse, after changes in school resources are controlled for. We also see that the share of private pupils is no longer positive, but negative (which is consistent with growth of low fee private schools of low quality).

Table 9 Learning premium from local density – Maths

	[1] Maths Coef.	t-stat	[2] Maths Coef.	t-stat	[3] Maths Coef.	t-stat	[4] Maths Coef.	t-stat
Ln Pupils in 2km radius	8.30**	[29.72]	2.01**	[4.98]	0.39	[0.77]	4.57	[1.44]
% Private in 2km radius					19.14**	[4.76]	40.99**	[3.04]
Pupil-teacher ratio			-0.82**	[-6.94]	-0.75**	[-6.20]	-0.41	[-1.62]
Multigrade			-5.16*	[-2.21]	-4.44+	[-1.89]	-8.55	[-1.64]
Single teacher			-3.14	[-1.10]	-4.24	[-1.49]	23.17*	[2.45]
Basic 5 Infra			1.87**	[3.42]	1.54**	[2.76]	-4.13**	[-4.47]
Ln Enrolment			13.12**	[9.21]	12.27**	[8.58]	23.34**	[5.31]
% Female			1.19	[0.32]	1.20	[0.32]	2.78	[0.38]
District years of schooling					1.03	[1.51]	-0.26	[-0.07]
District home quality					0.63	[0.40]	-16.85**	[-4.92]
Year 2016	91.92**	[74.22]	91.56**	[57.70]	90.11**	[52.96]	115.58**	[34.43]
Constant	441.48**	[240.94]	434.95**	[69.73]	437.53**	[65.46]	409.28**	[10.92]
School fixed effects	N		N		N		Y	
Observations	18,663		16,887		16,887		16,887	
R-squared	0.27		0.32		0.32		0.85	

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. Panel data with two years, 2007 and 2016.

Similar estimations are carried out for mathematics in Table 9. Here columns 1-3 show even quicker dilution of the density premium as more controls are added to the model. While the density premium remains positive even in the school fixed effects model, it is not statistically significant. Interestingly, in contrast with the results for reading, the presence of local private schools has a fairly large and positive coefficient in the final column; should the share of private schools increase by 10 percentage points, mathematics scores would increase by 0.1*40.99 = 4.1 points. Given the other controls in the models, this may be indicative of benefits of competition. Why this would take place in mathematics and not in reading, is puzzling.

Overall, the evidence from the OLS models without fixed effects indicates that the 'urban learning premium' can largely be explained by school resources and local wealth and education.

4.4 Evidence from Oaxaca decompositions

This Section provides an alternative approach to analysing the urban rural learning differences and their determinants. We divide districts into two groups on the basis of the level of urbanisation in 2007: those with an urbanisation rate over 80%, and those below. With these two groups, we rely on the Blinder-Oaxaca decomposition technique to assess which factors contribute to the educational differences between these two groups. We carry out the analysis separately for 2007 and 2016, which allows us to assess whether the nature and determinants of the disparities has changed. In practice, we use the common threefold decomposition, where the expected learning differences of the rural (Y_R) and urban municipalities (Y_U) are attributed to differences in endowments, or the differences in the levels of the explanatory variables (*E*), the differences in coefficients, or the marginal effects of explanatory variables (*C*), and the interaction of the two previous effects (*I*).[3]

Equation (2) states that the learning outcomes in both areas is a function of the *X* vector and the associated scalar of estimated coefficients (β). Equation (3) displays the standard decomposition of the differences between the predicted outcomes to the three components. The endowment effects, $[E(X_R) - E(X_U)]*\beta_U$, show how much lower the urban learning outcomes would be, should the urban schools have the same levels of endowments as the rural ones. The coefficient effect, $E(X_U)(\beta_R - \beta_U)$, measures the predicted change in urban schools if they had the estimated coefficients of the rural schools. The final term is an interaction of the two effects. The last term has a less obvious intuitive interpretation.

(2) $Y_i = X_i\beta_i + \varepsilon_i\ , \quad i = \{R,U\}$

(3) $$\begin{aligned} R &= E(Y_R) - E(Y_U) \\ &= [E(X_R) - E(X_U)]*\beta_U + E(X_U)(\beta_R - \beta_U) - [E(X_R) - E(X_U)]*(\beta_R - \beta_U) \\ &= E + C + I \end{aligned}$$

Table 10 presents the results. Comparing the overall difference between 2007 and 2016, we find that the urban districts had an advantage of 53.5 points in 2007, which had shrunk slightly to 43.6 by 2016. As explanatory variables, we use the same set of school-level control variables as in section 4.2, and for household characteristics, we use district-specific averages for years of schooling, and an index of four items that measure the quality of housing (water, electricity, toilet and sufficient space) from the Census. The numbers for 2016 are taken from 2017 census.

In terms of results, in both 2007 and 2016, the rows for the 'Total' variation suggest that the endowments explain nearly all differences between rural and urban districts. In fact, in both years, the decomposition suggests that had the urban districts had the same levels of explanatory variables as the rural districts, they would do even slightly worse than the rural districts. The overall effects from 'Coefficients' and 'Interaction' are small in comparison.

Looking at specific items in the production function for learning, the decomposition for 2007 suggest that the largest items that create the urban-rural gap are the local years of schooling (25 points), local home quality index (13 points), school size (11 points) and multigrade teaching (11 points). These factors highlight why urban-rural learning gaps are persistent – they are to a great extent down to factors that can't be changed with educational policy. The socioeconomic status of local parents can't be changed quickly, and small school size and multigrade teaching are hard to eradicate in sparsely populated areas. In 2016, the list of most important items is the same, although in somewhat different order.

[3] Eg. Oaxaca (1973), Blinder (1973), Jann (2008).

Table 10 Oaxaca decompositions of the urban district learning premium, 2007 and 2016.

Threefold Blinder-Oaxaca decomposition: Reading score in 2007						
				Observations:		
Rural Districts	446.06**	[643.97]		17,975		
Urban Districts	499.60**	[505.46]				
Difference	**-53.54****	[-44.36]				
	Endowments		Coefficients		Interaction	
Pupil-teacher ratio	0.09	[1.40]	-12.22*	[-2.44]	0.16+	[1.73]
Multigrade	-10.53**	[-6.09]	3.91**	[3.56]	8.22**	[3.57]
Single teacher	-3.31**	[-4.19]	0.44+	[1.82]	1.64+	[1.84]
Basic 5 Infra	-9.62**	[-9.49]	-7.69**	[-2.89]	3.65**	[2.88]
Ln Enrolment	-10.58**	[-5.08]	1.45	[0.10]	-0.33	[-0.10]
% Female	0.08	[1.48]	-5.97*	[-1.99]	-0.06	[-1.31]
District years of schooling	-24.95**	[-7.22]	-39.50**	[-3.70]	15.55**	[3.70]
District home quality	-13.01**	[-3.26]	26.71**	[2.69]	-12.11**	[-2.69]
Total	**-71.83****	[-27.90]	**1.58**	[0.62]	**16.71****	[4.89]
Threefold Blinder-Oaxaca decomposition: Reading score in 2016						
Rural Districts	542.80**	[773.00]		Observations: 9,969		
Urban Districts	586.42**	[721.26]				
Difference	**-43.61****	[-40.60]				
	Endowments		Coefficients		Interaction	
Pupil-teacher ratio	-2.19**	[-3.21]	-35.41**	[-7.85]	7.53**	[7.63]
Multigrade	-7.70**	[-5.87]	1.43**	[2.80]	4.62**	[2.82]
Single teacher	-1.75**	[-4.38]	0.41**	[3.63]	1.94**	[4.36]
Basic 5 Infra	-5.12**	[-5.44]	5.14	[0.88]	-1.01	[-0.88]
Ln Enrolment	-13.90**	[-8.31]	23.27*	[2.06]	-5.33*	[-2.06]
% Female	-0.07	[-1.10]	-4.34	[-1.42]	0.03	[0.90]
District years of schooling	-12.13**	[-5.05]	37.50**	[4.03]	-13.19**	[-4.02]
District home quality	-9.52**	[-3.49]	-40.74**	[-3.21]	9.76**	[3.21]
Total	**-52.38****	[-23.75]	**4.43+**	[1.95]	**4.34**	[1.46]

Notes: : **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets.

The estimated coefficients give an indication on the extent to which the rural-urban gaps in learning could be made smaller. The most obvious targets for policy are basic school infrastructure, and a reduction in multigrade teaching. These can be addressed with more money. School size appears to be highly important for learning as larger school have better outcomes, but creating larger schools in rural areas would imply difficult trade-offs with respect to school access. Household characteristics matter greatly for learning, but are obviously not subject to manipulation by short or medium term educational policies.

However, there is no evidence in this decomposition that the home resources would have become a larger determinant of learning over time, as for example the size of the endowment effects from district schooling levels and home environment quality has become somewhat smaller by 2016.

The two sets of analysis in this section do not entirely settle the question of why the major regions have not converged in learning outcomes, despite the fact that school resources have converged. The reform period favoured the more remote areas, resulting in substantial catch-up in basic school infrastructure and more favourable pupil-teacher ratios. Despite this, the catch-up in learning outcomes has not taken place. The persistent sources of regional inequality are the differences in household socioeconomic status, but also the larger average school size which can help the more urban areas maintain their advantage over rural areas.

5 Role of rural-urban migration in national learning outcomes

The main aim of this section is to focus on the impact of urbanisation from the perspective of an individual, and highlight a somewhat overlooked aspect of the relationship learning and urbanisation in Peru. The previous section has shown that learning disparities across the country appear to be to a substantial extent explained by differences in school resources, inputs and wealth across areas. One of the natural conclusions that follows, is that regional disparities could be relieved by investments in school resources and income across least developed parts of the country.

However, this is not the only possible way for students to gain access to an environment that is more conducive to learning. The other route, exercised by families, is migration to a more urban area where children can enjoy better school resources, have more choice in education, and more intellectual stimulation.

Migration matters also for another reason. The improvements in national learning levels in Peru have been impressive and have gathered international attention among educational policymakers. However, our period of study also coincided with rapid urbanisation, leading to shrinking pupil populations in rural areas, and increases in urban areas. It is therefore worthwhile studying the extent to which gains in learning were simply due to pupils moving to more developed locations.

This question is interesting academically, but also in terms of strategic policy formulation; the educational benefits of urbanisation are not yet fully known. The analysis also matters for evaluation of past policies: how much can educational policies and governments ‘take credit’ for improvements in learning, if the improvements are driven by rural-urban migration?

5.1 Decomposition analysis

Table 1 showed that many regions have experienced very significant levels of urbanisation.[4] For example, the share of public school pupils in Apurimac went from 48% urban to 70% urban between 2007 and 2016. Similar drastic changes have taken place in Cusco (54% to 75%) and Huancavelica (30% to 62%), and on a somewhat smaller scale in a number of other regions. There are limits to studying urbanisation, and in particular urban-rural migration with other Peruvian data sources. For instance, there is limited information on migration in the annual household survey and the census, available for 2007 and 2016 does not include much information on time of migration either. They are also representative at the region and district

[4] The numbers are computed from the School Census. As such, these are not directly comparable to census-based rates of urbanisation.

levels. Interestingly, the Peruvian school census can provide an annual picture of urban growth, at least with respect to the pupil population at a rather disaggregate level.

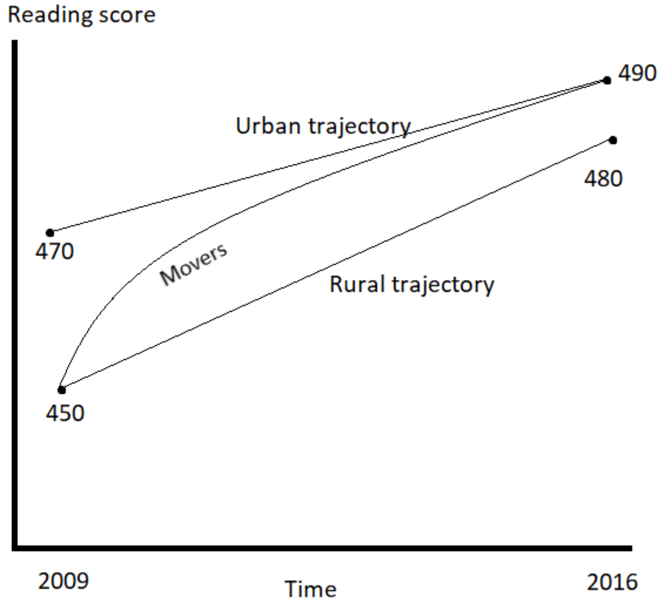
Table 1 also suggests that some of the largest improvements in learning outcomes have taken place in regions with rapid urbanisation. This highlights that the matching of pupils to better school resources happens in two ways: firstly by government action – by improving existing schools and provision in underprovided areas, but secondly, also by families moving to areas of better school quality. In fact, migration makes it easier to improve learning outcomes; it is easier to find teachers for urban schools, and it is cheaper to connect schools to water, electricity and internet in urban and peri-urban areas than rural ones.

In this subsection, we conduct a simple decomposition exercise to assess the extent to which the aggregate improvement in basic skills of second grade pupils between 2009 to 2016 is attributable to improvements in (1) urban learning outcomes, (2) rural learning outcomes and (3) movement of pupils from rural to urban areas. The purpose is to assess the degree to which the improvement is due to parent-led migration rather than policy to improve resources.[5]

We make the assumption that as pupils move from rural to urban areas within a region, their learning adjusts to the level of average learning in the urban schools. With basic skills such as those tested in grade 2, this is not a wild assumption. Consider a concrete example presented in Figure 21. Suppose that in a particular region, we find that from 2009 to 2016, average test scores in rural schools improve from 450 to 480. At the same time the urban schools improve from 470 to 490. Then assume, for example, that the share of second grade pupils that were urban has increased from 50% to 70% over the period. We now assume that this 20% increase in urbanisation benefited the region by moving a fraction of pupils to the average score of 490 instead of 480, or a 40 point increase on the starting point of 450. The 30% remaining in rural status, improve by 30 points, while those 50% who always remained urban, improved by 20 points. As such, had the additional urbanisation not happened, there would be a smaller improvement for the region as a whole. Had urbanisation remained at 50%, the overall regional improvement would have been 0.5*30+0.5*20 = 25. With 20% moving to urban area, the regional improvement to be 0.3*30+0.5*20+0.2*40 = 27. Thus, this simplistic calculation says that 2/27, or 7.4% of learning improvement would have been due to migration.

[5] Relative size of urban pupil population can also change due to differential birth rates, but this is likely to be quite small compared to migration.

Figure 21 Hypothetical trajectories for urban and rural schools in a region.



It should be recognised that this is simplistic, as it assumes that migration is not selective and the benefits are homogeneous. We also assume that the trajectories seen in the data are independent of migration. Furthermore, we assume that rural-urban population balance is due to migration and not differential fertility, even though this assumption does not really invalidate the exercise in itself. However, we still believe this exercise to be of interest since urbanisation in Peru has been rapid, and urban-rural differences in school resources can be very large.

To estimate these improvements from the data, it is helpful to give these quantities their own parameters. For reading scores, let: γ = the initial level rural schools, α = Rate of improvement in rural schools, β = the initial urban advantage, and δ = The difference between urban and rural improvements. In the example above, these parameters would be $\gamma = 450$, $\alpha = 30$, $\beta = 20$ and $\delta = -10$.

Assuming that the initial share of urban population in 2009 is U_{09}, then with these parameters, the total improvement without mobility of pupils would be

(4) $\alpha(1 - U_{09}) + (\alpha + \delta)U_{09}$

If we denote the share of movers as (U_{16} - U_{09}), the improvement in reading scores with mobility on the other hand, would be

(5) $\alpha(1 - U_{09}) + (\alpha + \delta)U_{09} + (\beta + \delta)(U_{16} - U_{09})$

The regional improvement in migration is (5)-(4), which simplifies to

(6) $(\beta + \delta)(U_{16} - U_{09})$

Equation 6 shows that the improvement has two components, the shift β to a higher urban learning trajectory, but also potentially slowdown in learning due to having more people in the urban area which improves more slowly (if δ is negative). Overall, the share of improvement due to mobility is [(6)-(5)]/(6)

To estimate these for all regions, the first step is to estimate urban and rural learning trajectories (as in Figure 21) for each region separately, using two cross sections of individual pupil scores:

$$(7)\ Read_{it} = \gamma + \alpha D_t + \beta Urban_{it} + \delta D_t * Urban_{it} + u_{it}$$

Here, t = (2009, 2016), $Read_{it}$ is the reading score of a pupil i, Urban is a dummy of whether i is an urban school, and D_t is a dummy for year 2016. Estimates provide the learning trajectories for rural areas (in 2009: $\hat{\gamma}$, and in 2016: $\hat{\gamma} + \hat{\alpha}$) and urban areas ($\hat{\gamma} + \hat{\beta}$ in 2009, growing to $\hat{\gamma} + \hat{\beta} + \hat{\alpha} + \hat{\delta}$ by 2016). Note that these trajectories will differ by region.

This specification abstracts from explicit effects of school resources because we want to account for all the factors that generate different results by area, whether due to school resources or other local factors. As such, in this calculation, pupils' learning trajectories are assumed to follow the path that average pupil in a region's average urban or rural area would follow.

This simple decomposition abstracts from selective migration and overlooks why urban areas might lead to different learning outcomes. As such, it is a simplistic back-of-the-envelope calculation. At the same time, it can be illuminating in showing us the degree to which pupil mobility has contributed to improved learning outcomes. One disadvantage is that this exercise can only be done for Spanish-language schools, as the bilingual schools are not tested in the second grade.

In Table 11, the total improvement (5) and component due to migration (6) are computed for all regions. The final column is the share of improvement attributable to 'movers', or more precisely, the increase in the share of urban secondgrade pupils. Lima and Callao are left out of the calculation due to being fully urban.

Table 11 Contribution of migration to improvement in learning between 2009-16

Department	Urbanisation in 2009	Change in Urbanisation	Total Improvement	Due to Migration	Share due to Migration
Amazonas	0.47	0.15	65	7	0.11
Ancash	0.53	0.23	57	16	0.28
Apurimac	0.51	0.25	73	19	0.26
Arequipa	0.81	0.14	56	8	0.14
Ayacucho	0.58	0.23	88	15	0.17
Cajamarca	0.31	0.23	49	11	0.23
Cusco	0.58	0.24	78	18	0.22
Huancavelica	0.35	0.31	64	15	0.23
Huanuco	0.36	0.26	61	15	0.25
Ica	0.85	0.11	56	5	0.09
Junin	0.45	0.38	54	22	0.41
La Libertad	0.53	0.20	46	14	0.31
Lambayeque	0.65	0.22	40	12	0.30
Loreto	0.57	0.17	59	10	0.17
Madre de Dios	0.79	0.08	69	3	0.05
Moquegua	0.77	0.21	65	12	0.18
Pasco	0.60	0.20	55	13	0.23
Piura	0.52	0.27	68	18	0.26
Puno	0.64	0.20	74	8	0.11
San Martin	0.58	0.15	70	9	0.13
Tacna	0.91	0.05	93	3	0.03
Tumbes	0.91	0.04	36	2	0.05
Ucayali	0.72	0.16	47	12	0.25

Notes: Sample is based on public Spanish-language schools only. Lima and Callao are excluded due to being fully urban. 'Total improvement' refers to increase in Reading score at grade 2 from 2009 to 2016. The final column 'Share due to migration' computes the estimated share of improvement due to migration as opposed to improvement of results in rural or urban schools.

Table 11 shows that in 12 regions (out of 23 examined), more than 20% the total improvement in reading scores is potentially due to migration, or the balance of pupils shifting towards urban schools. While this is a simplistic calculation, it communicates a powerful message: migration matters for aggregate results when urban public schools and environments are better resourced than their rural counterparts. Movement of people to urban areas aids the educational goals of the country by pairing pupils with better school resources, teaching and economic environments.

There will be second-order effects to this migration, some of which are positive and some negative. On the positive side, the inflexibility of teacher labour markets and migration automatically lead to improving pupil-teacher ratios in the rural areas, at least temporarily. In the longer term, rural depopulation may however reduce the political weight of rural areas further, risking the educational provision of remote areas in the longer run.

An obvious criticism against the above calculations is that there might be selective migration, i.e the families that move from rural to urban areas are more ambitious and academically oriented, and as such the calculation above will overestimate the effect of moving people to urban areas. This is possible, but the opposite may also hold: that many of the movers consist of disadvantaged people from the rural areas. Migration is always selective, but here one should keep in mind that the test scores examined in this exercise come from second grade pupils, and they focus on fairly basic skills in literacy and numeracy. These skills are possible for nearly everyone to master, irrespective of background.

5.2 Panel data evidence on rural-urban movers

The Department of Education in Peru has linked a subset of pupils' scores so that the pupils can be followed between grades 2 and 8. The first observation is thus during the second grade of primary school and the second observation during the second grade of secondary school. This panel covers pupils who are in the second grade of primary school in 2009, 2010, 2012 and 2013, and consequently in 8^{th} grade in 2015, 2016, 2018 and 2019.

The sample is not representative of all pupils. It is biased towards urban areas, and there is attrition in the sense that not all pupils who move between grades 2 and 8 are reached. In the panel, most observed moves are within the same region. Despite these shortcomings, the panel data allows us to study the value-added learning between grades 2 and 8, and for the current question, it allows us to study whether pupils who move from rural to urban areas, improve their learning more than those who keep attending a rural secondary school. This allows to examine in an alternative way, whether urbanisation is component in the improvement of learning outcomes in Peru.

Table 12 summarises the data. We describe two samples: the first sample consists of all pupils for whom panel data is available, and who either remain in rural or urban area, or move from rural to urban area. The second sample consists of pupils who are in rural primary school in the 1^{st} wave of observation, which is in years 2009, 2010, 2012 or 2013 depending on the cohort. We have excluded the urban-rural movers from the sample.

The summary statistics show that in the full sample, only 17% of pupils are in a rural primary school, 30% are in a private primary school, and 28% in a private secondary school. Among the smaller sample of rural-origin pupils, the majority, (60%) attend an urban secondary school. It may be that this does not necessarily involve moving house, as one may simply need to attend a secondary school in a town or a city. In the sample, 95% speak Spanish as native language, showing that the sample is biased towards cities (in Census 2017, 84% report Spanish as their native language).

Table 12 Summary statistics for value added between primary and secondary schools for rural-urban movers.

Sample:	All (n = 1,250,648)		Rural origin (n = 214,999)	
	Mean	S.D.	Mean	S.D.
Primary reading score	561.7	83.0	508.1	77.3
Secondary reading score	580.6	69.7	534.7	61.9
Value added reading	18.9	68.5	26.6	73.4
Primary maths score	552.8	107.0	511.6	103.1
Secondary maths score	574.6	85.7	531.1	73.4
Value added maths	21.9	93.2	19.5	100.1
Female	0.507	0.500	0.499	0.500
Native Spanish speaker	0.949	0.219	0.831	0.375
In rural primary school	0.172	0.377	1	0
Rural-Urban mover	0.104	0.305	0.605	0.489
In private primary school	0.307	0.461	0.043	0.202
In private secondary school	0.281	0.449	0.054	0.226
Cohort (1-4)	2.600	1.099	2.468	1.139

Notes: The panel data is not a random sample of Peruvian schools and may not be able to fully track all moving pupils, especially across regions. Data is shown only for pupils whose location is known for both primary and secondary school. We have excluded pupils who move from urban to rural areas. Cohort 1-4 refer to pupils who were in second grade in 2009, 2010, 2012 and 2013, and in 8th grade in 2015, 2016, 2018 and 2019.

In the next table, (Table 13), we estimate a simple value-added model, in which the dependent variable is the change in the learning outcome scores between the primary and secondary schools. Since the tests are normalised, the 'Value added' is typically close to zero as expected. This means that negative values imply that the pupil has fallen back in the national distribution, and positive values mean that the pupil has gained in his/her relative position.

We explain the value added with district of origin fixed effects, cohort effects, and whether the pupil swapped to an urban school between primary and secondary schools. The two alternative samples provide two alternative comparison groups. In the full sample, the movers are compared to all pupils who stay in their urban/rural category, and in the rural origin sample, the movers are compared to pupils who stay in rural secondary schools. The results are reported for Reading and Mathematics separately in panels A and B.

Across the specifications, value added in reading is 2.4-4.5 points larger for those who move from a rural primary to urban secondary, compared to comparison groups. Comparing it to the standard deviation of reading score in the secondary schools (69.7), this corresponds to 0.034-0.065 standard deviations. In Mathematics, the corresponding effect size is about 0.015-0.027 standard deviations. Overall, these effects, while positive, are relatively small.

Table 13 Rural-urban migration and value-added learning between primary and secondary schools

Panel A: Value-added in Reading				
	(1)	(2)	(3)	(4)
Sample	All	Rural origin	All	Rural origin
Rural-urban mover	4.163**	3.026**	4.546**	2.418**
	[0.224]	[0.373]	[0.240]	[0.382]
Private schools	Included	Included	Excluded	Excluded
Observations	1,250,684	214,999	798,892	199,907
R-squared	0.037	0.067	0.051	0.069

Panel B: Value-added in Maths				
	(1)	(2)	(3)	(4)
Sample	All	Rural origin	All	Rural origin
Rural-urban mover	1.346**	2.345**	2.067**	1.574**
	[0.305]	[0.508]	[0.327]	[0.521]
Private schools	Included	Included	Excluded	Excluded
Observations	1,250,199	214,997	798,711	199,903
R-squared	0.039	0.070	0.054	0.072

Notes: **: p <.01, *: p <.05, +: p <.10. Standard errors in brackets. All models control for cohort and district of origin effects

In Table 14 the results are broken down by native language and gender. We use the specification which has only rural-origin pupils, and include private schools. The results are again separately for Reading and Mathematics in panels A and B. The results show that the benefits of attending urban secondary are much larger for native language speakers. Surprisingly, this effect is even heightened in Mathematics, which should be more neutral to language. Spanish-language pupils get only marginal benefit from moving from rural to urban area, whereas native boys improve their score by 10.1 point (0.12 SD), and girls by 6.6 (.08 SD).

A direct comparison with the earlier decomposition is not possible since the decomposition only used test scores from the second grade, where there’s a larger gap between rural and urban results. In the panel data, the outcome variable, the value added, measures the (relative) learning between grades 2 and 8, and is based on a highly selected sample. The panel estimations here might underestimate the positive effect of moving, since we found these effects to be larger for native language pupils, who are underrepresented in the sample.

Overall, both sets of analysis in the section suggest that the recent movement of pupils to urban areas have had an effect on learning outcomes. For the basic literacy learned by grade 2, these effects can be very large – up to 30-40% of the overall improvement in some of the rapidly urbanising regions such as Junin, La Libertad and Lambayeque, and around 20% in many other regions. The panel data analysis of value-added scores between second and eighth grade point to much smaller effects, but these figures can’t be directly compared.

Table 14 Rural-urban migration and value-added learning, heterogeneity by language and sex

Panel A: Value-added in Reading

	(1)	(2)	(3)	(4)
Sample	Rural origin	Rural origin	Rural origin	Rural origin
Language, Sex{M/F}	Native lang, M	Native lang, F	Spanish, M	Spanish, F
Rural-urban mover	7.124**	6.793**	2.215**	1.748**
	[1.199]	[1.219]	[0.592]	[0.594]
Private schools	Included	Included	Included	Included
Observations	18,297	18,074	89,513	89,114
R-squared	0.145	0.146	0.072	0.075

Panel B: Value-added in Maths

	(1)	(2)	(3)	(4)
Sample	Rural origin	Rural origin	Rural origin	Rural origin
Language, Sex{M/F}	Native lang, M	Native lang, F	Spanish, M	Spanish, F
Rural-urban mover	10.115**	6.608**	0.858	1.343+
	[1.643]	[1.650]	[0.806]	[0.806]
Private schools	Included	Included	Included	Included
Observations	18,300	18,073	89,507	89,116
R-squared	0.135	0.139	0.077	0.082

Notes: **: p <.01, *: p <.05, +: p <.10. Standard errors in brackets. All models control for cohort and district of origin effects

6 Conclusion

Over the examined period, 2006-2016, the state made commendable improvements to primary education (as well as other stages of education). The funding increased significantly both in current and real terms. The data shows substantial improvements in the quality of basic school infrastructure. The period also saw numerous improvements in teachers' incentives, hiring and salaries. Many of the reforms aimed to improve education especially in rural and remote areas.

Given the nature of the reforms, it can be somewhat surprising that the overall regional inequality as well as inequality between urban and rural areas persisted during the whole period we examined. Urban public schools experienced rapid improvement in test scores, despite not being the main focus of the reforms or increased funding, and despite being stretched by rural-urban migration.

Our analysis does not suggest a simple explanation for the lack of regional educational catch-up, but that there are a range of factors. The prominence of private schooling in urban areas does not appear to explain the persistence in learning gaps.

We analyse the urban learning premium in learning in Peru over the period using regression analysis and Oaxaca-Blinder decompositions. These methods, while useful, are still constrained by the fact that improvements in the factors of the educational production function are correlated, meaning that it is impossible to precisely separate the impact that different factors have. Despite this, it can be said that roughly half of the educational achievement gap between urban and rural areas in the second grade of primary school, can be attributed to school resources such as basic infrastructure, teacher inputs and teaching arrangements such as multigrade teaching. The remaining part of the learning gap can be explained by local level of socioeconomic development, and importantly, the average size of schools. The data universally support the view that learning outcomes are on average better in larger primary schools, even after controlling for teacher and infrastructure inputs. This is likely to be due to better organisation and management, and possibly the (unobservable) quality of instruction.

Our results imply that the persistence of regional inequalities is to a significant extent due to factors beyond educational policy. At the same time, had the reforms not emphasised rural and remote areas with substantial regional catch-up in school resources, the regional disparities would have grown even wider than they were by the end of the examined period.

Finally, we examined a somewhat overlooked aspect of the improvement of school resources. In 2009, 63% of second grade Spanish-medium public primary school pupils were in urban schools. By 2016 this had grown to 82%. This represents a significant movement of families towards more urban locations, where schools typically have better levels of infrastructure and less difficulties in finding competent teachers. As show by our analysis, urban areas are also likely to have other positive effects on learning. We provide a simple estimate of the proportion of the learning gains between 2009-2016 that arises from the movement of the pupil population towards urban areas. We find that in around half of Peruvian regions, migration to urban areas contributes towards more than 20% of the learning gains. With pupil level panel data, we also find that the value added in reading and mathematics is slightly larger for those who move from a rural primary to urban secondary, compared to comparison groups. These results highlight the fact that educational progress is not only driven by additional resources, but also by the mobility of people.

7 References

Balarin, Maria (2021). The Political Economy of Education Reforms in Peru. Grade, Lima, Peru.

Blinder, Alan S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 8(4): 436-455.

Chetty, Raj, and Nathaniel Hendren (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics* 133(3): 1107-1162.

Jann, Ben (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4): 453–479

van Maarseveen, R. (2021a). The urban–rural education gap: do cities indeed make us smarter?. *Journal of Economic Geography*, *21*(5), 683-714.

van Maarseveen, R. (2021b). The effect of urban migration on educational attainment: evidence from Africa. *Available at SSRN 3836097*.

Majerowicz, S. (2015). Evaluaciones rápidas de impacto: acompañamiento pedagógico, soporte pedagógico intercultural (ASPI) y soporte pedagógico. *Reporte final. Lima: Grade/Forge*.

Majerowicz and Montero (2022). Can Teaching be Taught? Experimental Evidence from a Teacher Coaching Program in Peru. Mimeo

Oaxaca, Ronald (1973). Male-female wage differentials in urban labor markets. *International economic review*, 14(3): 693-709.

Saavedra, Jaime, and Marcela Gutierrez. Peru: A wholesale reform fueled by an obsession with learning and equity. In *Audacious education purposes: How governments transform the goals of education systems* (2020) (Ed. Fernando M Reimers), pp. 153-180, Springer Open.