The Urban Learning Premium - Evidence from Peru

Sonja Fagernäs[1], Panu Pelkonen[1,2], Juan Manuel del Pozo Segura[1,3]

17 April 2024

Abstract

Persistent regional inequalities in education and rapid urbanisation are common features in emerging economies. We examine the urban learning or schooling premium in Peru using three approaches: 1) estimating the effect of local population density on learning with a register of primary school pupils, 2) studying changes in learning with a panel sample of rural-urban movers, and 3) census-based estimations on the effect of the duration of urban exposure during childhood on school attainment. The degree of causality in the set-ups varies, but all methods confirm that urban areas produce higher learning or educational outcomes. The results suggest that the unconditional urban premium is largely explained by the socio-economic status of pupils' households and school characteristics. Analysis of pupils who move shows that those who move from rural to urban areas gain more than others between primary and secondary school, the effect being larger with for some disadvantaged groups. Sibling comparisons with census data further show that timely progression in the school system is not driven by selection or family effects, but rather, the time spent in urban areas. Overall, the results suggest that the ongoing urbanisation within developing and emerging economies is likely to provide a mechanical boost to aggregate learning outcomes by feeding children into better school environments.

1 – Dept. of Economics, University of Sussex

2 – IZA

3 – National Foundation for Educational Research (NFER)

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

This research has been funded by the British Academy.

1 Introduction

Rapid urbanisation is a common phenomenon across many low and middle-income countries, as is rural-urban inequality in the quality of schooling and learning outcomes. Nevertheless, the precise impact of urbanisation on learning and educational attainment remains understudied.

In this study, we rely on several methods and data sets to establish the presence and magnitude of an urban learning or schooling premium in Peru between 2007-2016. We also provide evidence of the potential drivers of the urban learning premium.

Peru is a relatively urban country, but many areas saw significant increases in the degree of urbanisation during the time period studied, on aggregate from 76% to 82% based on the population census, but the share of primary school children in an urban school increased more, from 70% to 86%. Some of the largest improvements in learning outcomes took place in regions that experienced rapid urbanisation. During the period studied, the education budget also increased substantially, and a range of educational reforms were introduced towards the end of the period.

Despite the relevance of urbanisation for economic development, there are surprisingly few studies that explicitly study how urbanisation affects learning and schooling with micro-level data, in particular in a developing country context. One connected literature in the developed country context is one that studies neighbourhood effects (see e.g. Aaronson, 1998, Potnick et al 1999, and Chetty and Hendren, 2018). For instance, using tax record data and sibling comparisons, Chetty and Hendren find that children who spend more time growing up in a better neighbourhood have better outcomes later in life, for instance in terms of earnings, college attendance rates and fertility. Another related literature studies international migrants, and their age of arrival to a new country (eg. Van Den Berg et al, 2014, and Basu 2018). However, in the case of international migration, the change in environment incorporates a broader range of factors, beyond urbanisation.

Van Maarseveen (2021a) relies on a similar methodological approach as in Chetty and Hendren (2018) and Alesina et al. (2021) to study the impact of the length of urban exposure for rural urban movers on educational attainment and choices. He shows that children growing up in urban regions in the Netherlands have higher levels of education than children in rural regions, controlling for cognitive ability. Focusing on urban areas in the UK, Gibbons and Silva (2008) study the effect of urban density on pupil test results in secondary schools.

In the developing country context, in another paper, van Maarseveen (2021b) studies how childhood urban exposure raises primary school completion, school attendance, and literacy rates in African countries, using population census data. As potential reasons for the urban education 'premium', he mentions higher returns to education in urban areas, lower travel costs, more choice and thus a better match between student and school, higher opportunity costs in urban areas and limited information on educational opportunities in rural areas. However, only suggestive evidence of the channels is provided. In the Peruvian context, a working paper by Cueto et al. (2019), using Young Lives data finds that migrating from a rural area to an urban area was associated with better learning outcomes, especially for under 8-year olds.

In this paper, we focus on differences on learning and school progression between urban and rural children in the context of urbanisation, using a range of data sets. This enables us to also provide evidence on some of the channels. Urbanisation can influence learning and schooling decisions via several channels, relating to the school system itself or independent of the school system. Urban schools often have better resources and better teachers. Urban schools are larger, which may lead to economies of scale, and can provide a more stimulating learning environment. Urban density alone may create competitive pressure on schools. Additionally, wealth and environmental differences between urban and rural households can matter for resources, demand for schooling and stimulation available outside school. Children's time use can also vary between urban and rural areas, with domestic or non-domestic duties being more frequent in rural areas. This in turn can affect time spent on homework, outside school classes or general alertness.

We focus on primary and secondary school aged children. We conduct three separate analyses; one estimating the association between nearby urban density and learning, and two focusing on rural-urban migration and its impact on learning and education indicators. Our test scores are primarily for second graders, but we also utilise a panel data set tracking children between grades 2 and 8, as well as population census data on school progression and attainment for children at both primary and secondary level.

For the first analysis, we create a measure of urban density within a 2km radius of each school ('micro-locality'), providing a more precise assessment of the degree of urbanisation compared to the simple urban-rural division. This is enabled by the geo-coding of schools.

Our estimates based on the school census and urban density in micro-localities provide evidence of a sizeable unconditional urban learning premium. On average, moving from the 10^{th} to the 90^{th} percentile in terms of density, increases reading scores by 58, which is about 70% of a standard deviation. This analysis also indicates that this urban premium can be explained by measurable household socioeconomic status and school characteristics. In the analysis, we highlight the differences between factors which are malleable by educational policy, and which are not.

More descriptive evidence indicates that while economic growth benefitted many areas over the period studied, the income gap between more urban and more rural areas remained high, and even increased as measured by monthly incomes and expenditures. We also demonstrate that the time-use and health of urban and rural children differ. In conclusion, while the urban schooling system itself can deliver better learning outcomes, the urban environment is more conducive to learning for several other reasons as well.

The analysis on the urban premium based on local density cannot fully control for selection of families across the different intensities of urbanization. For a more causal interpretation, we rely on two further methods to establish an urban learning or schooling premium, by focusing on rural-urban migrants. Firstly, with a panel data set tracking 4 cohorts of pupils between the second and eighth grades, we find that 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. Secondly, we rely on the Peruvian population census and a sample restricted to 7-18 year old children, whose families had moved from rural to urban areas. The identification approach is similar to that in van Maarseveen (2021b), adapted from Chetty and Hendren (2018). We find that teenagers, who had moved to an urban area earlier in their life have higher school attainment, measured using a number of indicators. In addition, sibling comparisons with the full range of school aged children and differential sibling exposure to urban schooling show that timely progression through the school system is affected by the time spent in an urban school environment. Since sibling comparisons account for family factors, this effect is not driven by selection, which highlights the importance of the urban environment itself and associated factors. Overall, our results suggest that differences in learning environments do not only manifest themselves in levels of learning, but also in timely starting of schooling and progression throughout childhood.

The paper is organised as follows. Section 2 describes the data. Section 3 provides descriptive statistics on developments in schooling and socio-economic indicators in urban and rural areas. Section 4 focuses on the association between urban density and learning, including an analysis of potential channels of effect. Section 5 focuses on the role of rural-urban migration in learning gains and the impact of urban exposure on broader schooling outcomes. Section 6 concludes.

2 Data

For data on schools, we use the primary school data from the Censo Escolar for 2007-2016, which is an annual school level census data set, which covers all schools in Peru. Among other things, the school census data include school level aggregates on resources, teachers, pupils and location, but does not contain information on household or parental characteristics. The data are geo-coded, so we know the precise location of each school. We link these data to annual data on test scores for second-grade pupils from the Evaluación Censal de Estudiantes (ECE) for the same years. This includes nationally comparable data for Reading and Mathematics[1]. Second graders are the group for whom test score data are available for the longest time period, 2007-2016, starting much earlier than for any other grades.

In addition, we utilise a large separate ECE sub-sample, which tracks approximately half of the pupils in the school census from grade 2 to 8 for four cohorts, who are in grade 2 in years 2009, 2010, 2012 and 2013. The tracked sample is not representative of all pupils. It covers slightly over half of the relevant cohorts and it is slightly tilted towards urban pupils (77% in panel vs 70% in school census in 2009)**.**

Additionally, for information on population and household characteristics we use 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. We also use the latter for the analysis on the impact of rural-urban migration on school progression and attainment of children and teenagers. The Census data are available for the entire population. Finally, we also use the fourth round (2013) of the Young Lives data set for some descriptive statistics on differences in time use and habits between urban and rural children. This is a longitudinal survey following a younger and older cohort of children. We use the data for the younger cohort, who are 12 years old at the time of the survey.

3 Persistence of rural-urban inequality: descriptive evidence

This Section provides a descriptive analysis of developments in urban-rural differences in schooling and learning as well as socioeconomic development over the time period in question. We categorize Peru's districts, which are sub-regional units, into urban and rural. Districts are defined as 'urban' if their rate of urbanisation was over 80% in the Census of 2007 (391 districts), and 'rural', if below 80% (1373 districts). The data relate to Spanish speaking schools, as bi-lingual schools, which are largely rural, test pupils at a different age.

Figure 1 indicate that second grade test scores have improved since 2007 in both urban and rural areas, but rural areas have not caught up with urban areas; in fact 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,

[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.

but rural areas continued to lag behind urban areas. The data on teacher qualifications in the School census is not comparable across the period studied, but as a potential indicator of teacher quality, we use an indicator for the share of teacher with a permanent versus a temporary contract.

The relative improvement in infrastructure in rural areas partly coincides with a substantial boost to financing since 2011, some of which was targeted at rural areas. The economic conditions of households and parental inputs are also likely to be important factors in learning, although more difficult to assess.

In the graph in the upper left side corner of Figure 2, a 'household basics' indicator based on the 2007 and 2017 population censuses is plotted against the degree of urbanisation at the district level in 2007. While urban districts had an advantage with respect to household quality both in 2007 and 2017, there was substantial catch-up in more rural districts.[2]

Basic household services such as electricity and toilets cover one dimension of household well-being and resources. The other two graphs in Figure 2 rely on data from the household survey ENAHO on average hourly pay, aggregated to the level of 25 regions between 2007-2016 and divided by urban and rural location. They show that urban areas experienced more rapid growth in pay and household expenditure than rural areas.

Finally, Table 1 compares time use pattern of 12-year olds in Peru based on the younger cohort of children in the Young Lives survey Wave 4. Time use varies between rural and urban areas, which can have potential implications for learning outcomes or test scores. Rural children take longer to travel to school, have more caring and domestic responsibilities, their school days are slightly shorter and time spent studying outside school is shorter. They have significantly lower height for age scores, reflecting a possibly different health environment as well.

Overall, the descriptive statistics indicate that an urban learning premium has persisted despite targeted investment in rural areas, and that the premium is likely to be explained by a combination of school level and socio-economics differences.

4 Estimating the urban learning premium in 'micro-localities'

In this Section, we provide estimates of the association between urban pupil density and learning with pupil level data. We focus on primary schools and second grade test scores. We rely on the

[2] 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.

full sample of primary schools to create a measure or urban density. However, instead of the full school census, for the estimation on learning, we rely on the panel dataset of a subset of primary school pupils in grade 2, who are tracked again in grade 8.[3] These data are available for 2009, 2010, 2012 and 2013. The main reason for this is that information on family socioeconomic status is recorded when the pupils are in grade 8. We rely on the values for individuals once they are in grade 8 to proxy for the family socioeconomic status in grade 2. While this is not ideal, socioeconomic status is measured as an index and it is possible that the relative rankings have not changed substantially between grades 2 and 8.

We proxy urban density of a school with the total number of primary school pupils that are enrolled in schools within 2km of pupil's own school itself. This measure varies quite dramatically across schools and is correlated with learning outcomes. Figure 3 displays a simple relationship between the average reading scores across this measure of local pupil density for 2016. The left-hand side image uses a linear scale in the x-axis, while the right-hand side uses the natural logarithm of the number of pupils within 2km. The left-hand side image shows that average learning outcomes increase rapidly and monotonously until the number of pupils reaches around 15000 with 2km radius. This would correspond to a dense urban area. The right-hand image shows that the relationship is nearly log-linear after the value of about 4, which corresponds to 55 pupils within the 2km radius.

We use the logarithm of the number of primary school pupils within a 2 km radius of each school as our key measure of density. The location of schools is available from the school census. The summary statistics for the variables used in the regressions are shown in Table 2. Table 2 suggests a mean socioeconomic status index of 0.162, which by definition is 0.16 SDs above the mean.

We estimate the following pupil-level model for both Reading and Mathematics scores:

$$(1)\ Score_{ist} = \alpha \ln(Density_{st}) + \boldsymbol{X_{it}\beta} + \boldsymbol{Z_{st}\theta} + \delta_t + \varepsilon_{it}$$

where t refers to years, i to individuals and s to schools. The pupil learning scores ($Score_{ist}$) are explained with a vector of pupil characteristics ($\mathbf{X_{it}\beta}$) and school level characteristics ($\mathbf{Z_{st}\theta}$). The model controls for year effects (δ_t). Pupil characteristics include the socio-economic index, based on the information in grade 8. In addition to the household socio-economic index, for each pupil we create variables for the average socio-economic index of the pupil's peers in the school, and

[3] The data set has a slight urban bias, compared to the general population. The key reason for this is that the second graders, who are not matched to their 8th grade data on socioeconomic status are missing. On the other hand, there is some opposite selection due to some private schools (which tend to be urban) being excluded from the sample due to not reporting their school resources.

the average socio-economic index within the 2km radius of a pupil's school, excluding the school itself.

Table 3 shows the results for the Reading scores and Table 4 those for the Mathematics scores. Column 1 in Table 3 establishes a simple correlation based on an OLS model; more dense locations have better learning outcomes. The correlation is strong; density and the year effects alone explain 9% of the variation in the reading scores. On average, moving from the 10^{th} to the 90^{th} percentile in terms of log density, increases reading scores by 10.78*(10-4.7) = 57, which is about 70% of a standard deviation (80 in the sample).

In column 2, we add basic individual control variables for gender, Spanish speaker, and home socioeconomic status. This reduces the density premium to less than half of the original, from 10.78 to 4.21. This suggest that more than half of the urban premium is due to selection of families, especially by socioeconomic status.

In the third column, school level characteristics are added. This reduces the density premium below zero, suggesting that the urban learning premium can be fully explained by selection of family and school characteristics. School variables most strongly associated with learning (according to t-statistics) are the socioeconomic status of the peers and the share of tenured teachers. Adding further characteristics of the local areas in column 4 does not affect the conclusion.

The estimates for Mathematics in Table 4 show an even quicker dilution of the density premium as more controls are added. The density premium is smaller to begin with, and nearly all of it is explained by individual pupil characteristics. Again, the school characteristics most strongly associated with learning are the socioeconomic status of peers, and the share of teachers who are tenured. Interestingly, for Mathematics (as well as Reading), schools that report being 'rural' have better learning outcomes, even after the local population density is controlled for.

Overall, the evidence from these OLS models indicates that the 'urban learning premium' is largely unconditional, but can be explained by individual and school characteristics.

5 The role of urban-rural migration in educational gains

For a more causal estimate of the urban learning premium, we focus on rural-urban migrants, an alternative avenue to studying whether urbanisation is a component in the improvement of learning outcomes in Peru.

Figure 4 plots the improvement in test scores by region against the regional rate of urbanisation, based on two separate data sources. We measure the change in urbanisation from the population census as well as from the school census in the form of the increase in the share of second grade pupils in urban schools. The increases in the share of urban pupils are often larger than the changes in the share of the region's population living in urban areas. This suggests that some pupils who reside in rural areas, might have started attending school in urban areas.[4] The graphs show a clear positive connection between urbanisation and Reading scores, and a positive, but slightly weaker connection between urbanisation and Mathematics scores.

This Section contains two separate analyses, based on different data sets. The first analysis relies on a value-added model for the sub-set of children in ECE who are tracked between the 2nd and 8th grade for 4 cohorts. While there is some selection, the panel dataset allows us to study the value-added in learning between grades 2 and 8, and whether pupils who move from rural to urban schools, improve their learning more than those who keep attending a rural secondary school. In this panel, most observed moves are within the same region.

In the second analysis, we use the Peruvian population censuses for 2007 and 2017 and compare educational outcomes of 16-18 year olds who moved to urban areas more recently to outcomes for those who have spent longer in urban areas. We also compare outcomes for siblings aged 7-18, who have spent a shorter versus a longer period of their schooling in an urban area.

5.1 Rural-urban migration and learning

Table 5 summarises the data, based on the sample of pupils who are tracked between grades 2 and 8. We describe two samples: the first sample consists of all pupils for whom data are available, and who either remain in rural or urban areas, or move from a rural to an urban area. The second sample consists only of pupils who are in a rural primary school in the first wave of observation, in years 2009, 2010, 2012 or 2013, depending on the cohort. We include pupils in both public and private schools.

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 their native language, showing that the sample is biased towards cities (in Census 2017, 84% report Spanish as their native language).

[4] An alternative possibility is that the school census has a different definition for urban locality than the national census. We have not been able to confirm whether this is the case.

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. It takes the following form

$$(2) \qquad \Delta Score_i = \alpha + \beta D_{mover} + \gamma_d + \delta_c + \varepsilon_i$$

We explain the value added with district of origin fixed effects (d), cohort effects (c), and whether the pupil swapped to an urban school between primary and secondary schools (D_{mover}). We rely on two alternative comparison groups. With the full sample, rural-urban migrants are compared to all pupils who stay in their urban/rural category. With a more limited sample, rural-urban movers are compared to pupils who stay in rural secondary schools. Since the test scores are normalised, the 'Value added' is typically close to zero as expected. This means that negative values imply that the pupil has fallen behind in the national distribution, and positive values mean that the pupil has gained in terms of relative position.

The results are reported for Reading and Mathematics separately in panels A and B of Table 6. Across the specifications, the value added in Reading is 2.4-4.5 points larger for those who move from a rural primary to an urban secondary school. Give that the standard deviation of the Reading score for secondary school is 69.7, this corresponds to 0.034-0.065 standard deviations. In Mathematics, the magnitude of the corresponding effect is about 0.015-0.027 standard deviations. Overall, these effects, while positive, are relatively small.

In Table 7 the results are broken down by native language and gender for the sample that is limited to rural-origin pupils, but includes both public and private schools. The results are again separated for Reading and Mathematics in panels A and B. The results show that the benefits of attending an urban secondary school 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 a marginal benefit from moving from a rural to an urban area, whereas native boys improve their score by 10.1 points (0.12 SD), and girls by 6.6 points (.08 SD).

It is worth noting that this panel data sample is more selected than the full sample of schools and pupils. The estimations here might underestimate the positive effect of moving, for instance since we found these effects to be larger for native language pupils, who are under-represented in the sample.

In interpreting the effect sizes, one must keep in mind that there is no information on when the pupils changed to an urban school. It is likely, that many pupils change school between primary and secondary, which would imply that the mobile pupils would have experience urban school for only about 1.5 years. However, some pupils may have moved earlier, during the grades 3-6 of primary school. As such, the precise 'treatment' in this estimation holds some uncertainty.

5.2 Rural-urban migration and school progression

Rural-urban migrants are a selected group of people, both along observable and unobservable dimensions. To reduce the effect of this selection in estimates, we rely on an approach from Chetty and Hendren (2018) to study neighbourhood effects in the US. Van Maarseveen (2021b) uses a similar idea in a development context to study the effects of an urban environment on educational outcomes. He uses African population censuses to compare secondary attainment of teenagers who have moved to an urban location earlier versus later. It is argued that while movers are in general a selected sample, there is little difference in selection between those who moved earlier versus those who moved later. Chetty and Hendren focus more strictly on a sibling comparison, to take family fixed effects into consideration.

We utilise the same idea with Peruvian population censuses for 2007 and 2017. These are the only censuses available for the time period that we study, but they include the entire population. We focus on those children or youth who have moved from rural to urban areas and conduct two separate pieces of analysis. Firstly, we compare educational outcomes for 16-18 year olds who moved to urban areas more recently to outcomes for those who have spent longer in urban areas. Secondly, we also compare outcomes for siblings aged 7-18, who have spent a shorter versus a longer period of their schooling in an urban area.

The Peruvian Census allows us to identify children in internal migrant families by comparing their district of birth to the current one. We therefore need to classify districts into urban and rural, based on a threshold of 50% of the population being urban versus rural, and conduct robustness checks relating to this threshold. The district of birth is defined based on the district in which the child's mother lived in when the child was born. If the current district of residence differs from the birth district, the child is identified as a migrant. The data also include information on the district of residence five years ago, which we use to identify whether children moved more than once.

Among the internal migrants, the 'treatment', or the time spent in an urban location is defined by an answer to the question, 'Did you live in a different district 5 years ago?'. If the household has moved within 5 years, the family is classified as a 'Recent mover' with a shorter exposure to an urban environment, while if they have not, they have a longer than 5-year exposure to the urban environment.

In the first analysis, we compare teenagers who have moved to urban areas across households. We focus on 16-18 ears olds and several outcomes of interest: whether the teen has studied beyond primary school, whether they are of the correct age for the grade that they attend, whether they have graduated from secondary school and whether they are currently studying beyond secondary school. In principle, secondary school has been compulsory in Peru since 1993, but nevertheless not everyone in the secondary age category attends secondary school.[5] In our sample of movers, 92% report ever having attended secondary school and 53% report having completed the full five years of secondary school. As not everyone reports the number of years attended, the sample for the latter indicator is smaller than for the former.

Table 8 shows the possible migration patterns of 16-18 year old rural origin teenagers who have moved since their birth. As shown in the Table, the "treatment" category is defined on the basis of longer residence in urban districts, both if the individuals have stayed in the same urban district. Those who moved more recently to an urban area are in the "Control" group. In the analysis, we thus compare sets 1 and 2 from Table 8. Sets 2 and 3 could in principle be used as well, but these are small samples and have moved twice, so we leave them out of the analysis.

We estimate the effect of duration of residence in an urban environment using the pooled sample of two censuses and the following model

$$(3)\ Education_{ic} = \alpha + \beta Treat_{ic} + \boldsymbol{X}_{ic}\theta + \pi_c + \sigma_a + \gamma_d + \varepsilon_{ic},$$

where *i* refers to individual, *c* to census, *a* to age and *d* to current district. Education refers to the set of different educational outcomes. The treated group are those who moved to an urban area more than 5 years ago, and the control group refers to those who moved to an urban area within the last 5 years. We expect those in the treated group to have better outcomes (positive β). The model controls for age dummies, census dummies and fixed effects for the current district of residence. We also control for several characteristics of mothers ($\mathbf{X}_{ic}$). By the age 16-18, the children typically have finished or are close to finishing their secondary schooling but are unlikely to have moved away from their parents.

Table 9 summarises the pooled data of urban internal migrants, pooled from the Censuses 2007 and 2017. Table 9 shows a z-test for the differences in the group means between the treatment and the control group. The key background variables, the maternal characteristics, are reasonably similar despite the non-randomised setting. However, mother's age and whether she has studied beyond primary education differ between the treatments significantly. In the outcome variables, the statistical differences between treatment and control are much stronger.

Results are presented in Table 10. Longer exposure to an urban location leads to a nearly 3 percentage points higher likelihood of studying beyond primary. It also leads to a 3 percentage points higher likelihood of attending the correct grade to age, and a percentage point higher

[5] The length of primary schooling is 6 years and secondary school is 5 years.

likelihood of graduating from secondary school. There is a small, marginally significant effect on being enrolled in studying beyond secondary school.

There may be several reasons why the educational outcomes are better in urban areas. Part of this may reflect the stronger test results seen in the previous analysis, but there may also be several direct explanations to the urban environment being more conducive to higher educational attainment, from better school availability, enforcement, higher returns to education to opportunity costs.

In our second analysis, we estimate a within-family model of exposure to urban areas. This relies on the fact that depending on their age, different siblings have spent a larger part of their time in education in an urban area. Suppose that a family has moved to an urban area 4 years ago with siblings aged 10 and 13. In this case, the younger sibling has spent all of her schooling years (assuming from age 6) in an urban area, while the older one started her schooling in the previous location.

We constrain the sample to urban dwellers who have lived in another district five years ago. We then create a variable indicating the age difference to the oldest sibling, 'AgeDif', which measures the relative exposure of younger siblings to the urban area and interact this with the rurality of the previous district of residence, based on the share of the district population that is rural (between 0-1). The following model is estimated:

(4) $$Education_{ic} = \alpha + \beta AgeDif_{ic} * OrigRurality_d + \delta AgeDif_{ic} + \theta Fem_{ic} + \sigma_a + \gamma_{family} + \varepsilon_{ic}.$$

Controls include gender, age effects and family fixed effects (which also cater for current location, origin and census year). Here, β measures the urban advantage attributed to being one year younger, when a family moves from a (fully) rural to an urban area. The appeal of the estimation lies in the ability to control for heterogeneity across families, and account for selection, given that migration is a choice. Parameter δ measures the disadvantage that younger siblings may otherwise have in terms of educational attainment.

The drawback of the approach is that there are only a few potential outcome variables in the Census data, given the age range of the siblings. Namely, we will estimate the effect of urban exposure on being in the correct grade to age. In the full sample, only 75% of pupils are in the correct grade to age. Failure to be in the correct grade may be due to starting school late, not being in school, or having repeated a grade. By law, all children should be in school until the end of the secondary school, so deviations should in principle be minimal. Another variable available to us would be whether the child is in school. Table 11 shows the summary statistics for the relevant sample, compared to all urban residents between the age of 7-18.

Table 12 displays the results with and without family fixed effects. The first column estimates equation 3 as it is, and the second column the family fixed effects are replaced with district of origin fixed effects. The results in the first column suggest that being an additional year younger than an older sibling implies a higher likelihood of being in the correct grade by roughly 2 percentage points if the origin is fully rural. This is close to the average female advantage, which we estimate to be 1.6 percentage points. The main effect of the 'exposure' or the age difference is generally positive, suggesting that younger siblings are more likely to be in the correct grade also in general.

In the second column of Table 12, the family fixed effects are removed. Interestingly the effect of urban exposure on timely school progression is about half smaller, even if it is positive and significant. Given that the family effects in column 1 adjusted for bias due to selection of families, one can conclude that the effect of the urban environment may be underestimated when family fixed effects is not available.

Table 13 provides two additional estimations. In column 1, we re-estimate equation (3), but estimate the effect of exposure to urban schooling non-parametrically, interacting each additional year of exposure. The estimates show that the effects of exposure grow nearly monotonically, but are statistically significant only after the siblings have 7 years of age difference. The second column estimates equation (3) using another outcome variable, whether the child is in school. Here too the effects are statistically significant and positive.

6 Conclusion

Across the world, there is relatively little systematic documentation on the impact of urban environment on learning and schooling, and the fundamental reasons behind it. This study provides a systematic analysis of the differences in educational outcomes of children across the rural-urban dimension, in an emerging economy of Peru.

We employ the Peruvian school census, household surveys and censuses from the period of 2007-2017. We conduct three complementary pieces of analysis. Firstly, we examine the relationship between geocoded urban density and test scores of primary school pupils, both unconditionally and including pupil and school controls. The key findings are that the urban learning premium is vast and increases monotonously with higher population density. Pupil-specific socioeconomic factors play a large role in explaining the premium, while the rest of the premium can be explained by school characteristics. This result suggests that school quality (including the socioeconomic status of the peer group) is a substantial mediator in the urban-rural learning differences.

The first analysis cannot rigorously account for selection of families, and as such, doesn't settle the question on selection versus resources. Our second analysis uses pupil level panel data on test scores and studies switching from rural to urban schools as pupils move between primary and secondary level. This analysis suggests that the value-added between grades 2 and 8 is larger for pupils who move from rural primary schools to urban secondary schools. The positive effects of urbanisation are larger for indigenous language speakers, for males and in mathematics as opposed to reading. The effects are relatively small. The key difference in relation to the urban density estimations is that family effects are better accounted for due to the value-added framework, but differences in school resources aren't controlled for due to differences between primary and secondary levels.

In the final analysis, we rely on population census data from 2007 and 2017, and construct the location history of families. We compare children who have moved from rural areas to urban areas, but with a different length of urban exposure and siblings who have spent a varying share of their time in school in more urban areas due to the migration patterns of the family. We find that for samples of movers, a longer exposure to urban districts leads to a higher likelihood of timely school progression in terms of age-grade relationship, starting school in time, and graduating from secondary school, and studying beyond secondary schooling. An analysis of sibling differences in urban exposure suggests that timely school progression in urban areas is not due to family selection, but attributable to the time spent in an urban area.

Overall, the analyses show that the urban learning or schooling premium in Peru is pervasive, and some of it persists even in analyses that account for family or pupil fixed effects. As such, it cannot be explained fully by selection. Local environments matter and the associated school resources shape the educational trajectories of pupils.

From a public policy perspective, one striking conclusion is that a non-trivial part of recent improvement in national learning outcomes in Peru, as witnessed in test scores in school census, but also in international assessments such as PISA, may be driven by Peru's rapid urbanisation. This result may carry to many other in developing and emerging countries; urbanisation can mechanically lead to improvements in national learning outcomes.

7 References

Aaronson, Daniel (1998). Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes, *Journal of Human Resources*, 33(4), 915-946.

Alberto, Alesina, Sebastian Hohmann, Stelios Michalopoulos, Elias Papaioannou (2021). Inter-generational Mobility in Africa. *Econometrica*, 89(1).

Basu, Sukanya (2018). Age-of-arrival effects on the education of immigrant children: A sibling study. *Journal of Family and Economic Issues*, *39*, 474-493.

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.

Cueto, Santiago, Claudia Felipe and Juan León (2019*)*. Venciendo la adversidad: trayectorias educativas de estudiantes pobres en zonas rurales del Perú (Documento de Investigación, 100). Lima: GRADE.

Gibbons, Stephen and Olmo Silva (2008). Urban density and pupil attainment. *Journal of Urban Economics*, *63*(2), 631-650.

Plotnick, Robert D., and Saul D. Hoffman (1999). The effect of neighborhood characteristics on young adult outcomes: Alternative estimates. *Social Science Quarterly*, 1-18.

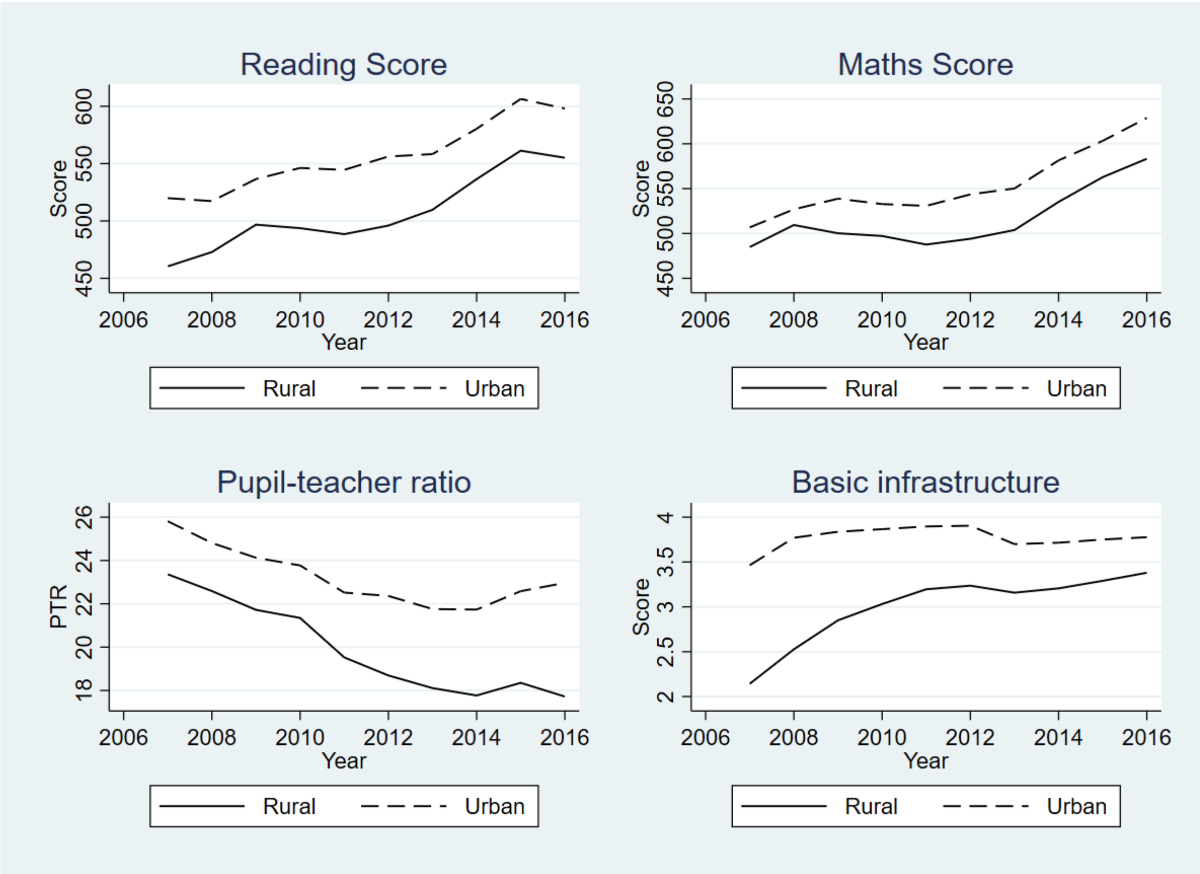
Van den Berg, Gerard J., Petter Lundborg, Paul Nystedt and Dan-Olof Rooth (2014). Critical periods during childhood and adolescence. *Journal of the European Economic Association*, *12*(6), 1521-1557.

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

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

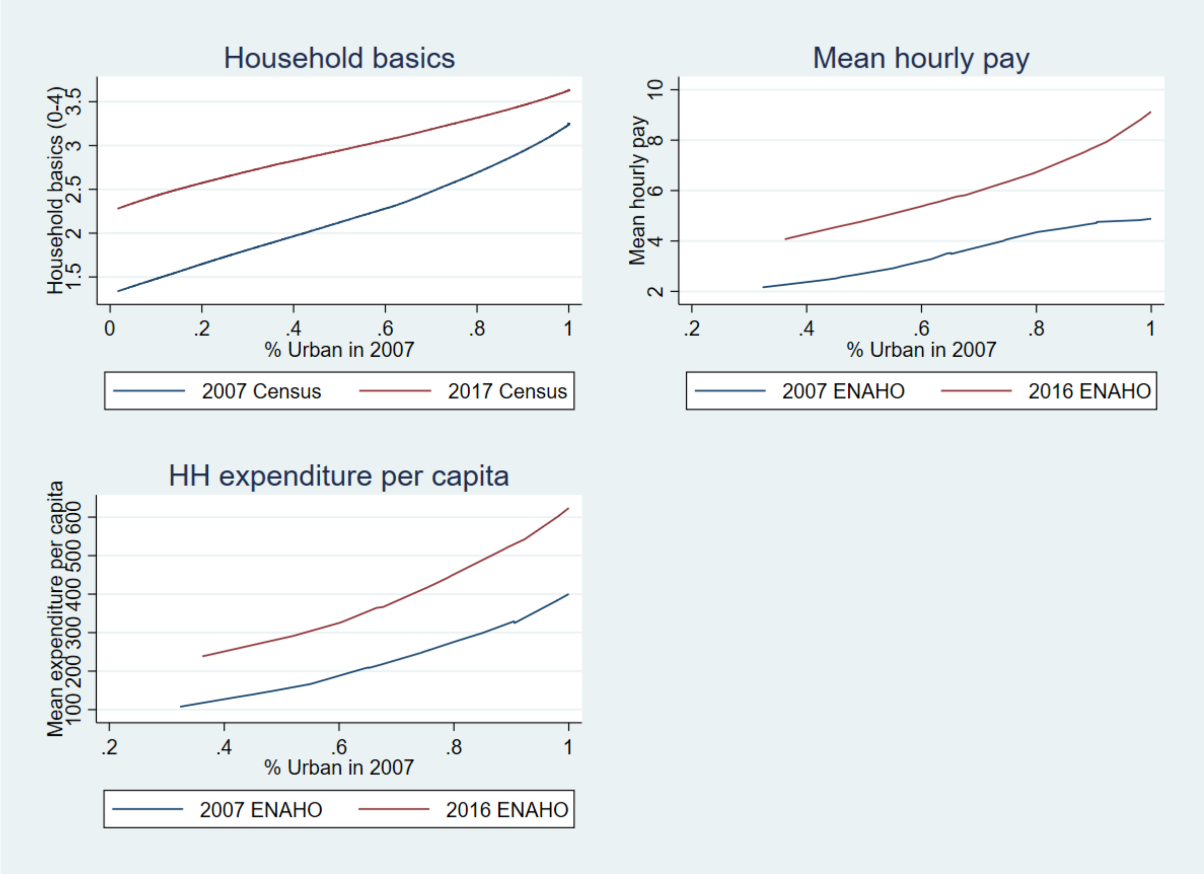
Figures and Tables

Figure 1 Second grade test scores and key school resources over 2007-2016 by urban/rural districts: primary schools



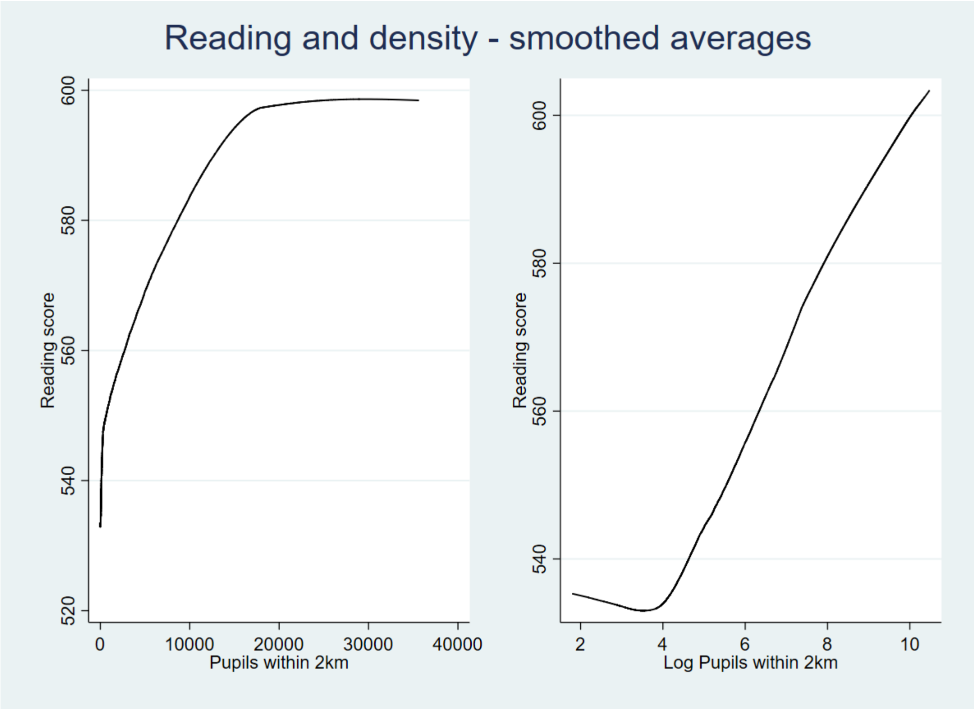
Notes: Districts are defined as 'urban' if their rate of urbanisation was over 80% in Census 2007 (391 districts), and 'rural', if below 80% (1373 districts). Basic infrastructure is a sum of four indicators (1-4) for which we have comparable data for the time period studies: Electricity, Water, Sewage and Toilet.

Figure 2 Catch-up in basic household infrastructure, hourly pay and expenditure, by initial level of urbanisation



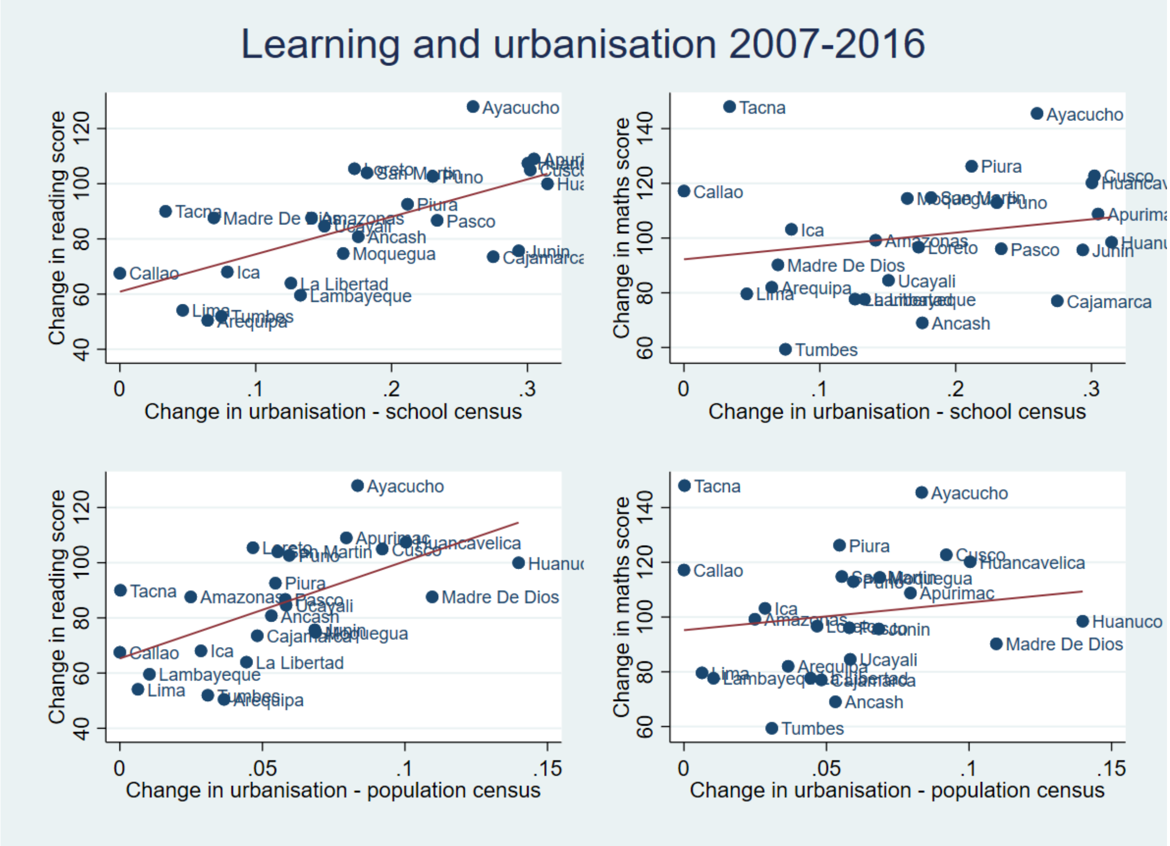
Notes: 'Household basics' is a district-level sum of 4 household-level indicator variables: Adequate water supply, Electricity, Toilet, and Non-overcrowding. Source, Censuses 2007 and 2016. Urbanisation measured by district. 'Hourly pay' is by region in 2006 and 2016 (Running smoothing). Data sources: ENAHO 2007 and 2016 for wages, Census 2007 for rate of urbanization by region. 'Mean household expenditure per capita' is by region in 2006 and 2016 (Running smoothing). Data sources: ENAHO 2007 and 2016 for expenditure, Census 2007 for rate of urbanization.

Figure 3 Relationship between 2km radius pupil density and Reading scores (second grade)



Notes: Based on full school census 2016 for primary schools and second grade test scores. Smoothed averages of school's average reading scores are plotted against measures of local geographic pupil density using stata's *lowess* smoother.

Figure 4 Change in urbanisation and 2nd grade learning scores across 25 departments/regions, 2007-16



Notes: Sample is based on public Spanish-language schools only in 2007 and 2016. Top images use the school census data to compute the change in urbanisation and refer to primary school pupils only. Bottom images compute the change in urbanisation from population censuses 2007 and 2017.

Table 1 Differences between urban and rural pupils in Peruvian Young Lives data

	Urban			Rural		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Travel time to school (min)	1,375	13.03	9.62	493	17.86	15.54
Hours of sleep	1,375	9.52	1.01	493	9.51	0.96
Care for others (hours)	1,375	0.78	1.03	493	0.92	0.95
Domestic tasks (hours)	1,375	1.15	0.74	493	1.39	0.70
Tasks on family farm (hours)	1,375	0.30	0.79	493	1.18	1.20
Activities of pay outside household (hours)	1,375	0.05	0.43	493	0.06	0.48
Studying in school (hours)	1,375	6.12	0.80	493	5.94	0.79
Studying outside school time (hours)	1,375	1.94	0.94	493	1.61	0.75
Playtime/Leisure (hours)	1,375	3.79	1.43	493	3.27	1.36
Height for age (z-score)	1,366	-0.74	2.64	485	-1.65	0.96

Notes: Young Lives Round 4, 12-year olds

Table 2 Summary statistics for public primary school pupils

N = 831,296	Mean	Std. Dev.	Min	Max
Year	2011.2	1.7	2009	2013
Pupil variables				
Reading score	563	80	112	814
Mathematics score	558	106	53	944
Socioec. Index	0.162	0.888	-3.480	9.436
Female	0.513	0.500	0	1
Spanish native language	0.957	0.203	0	1
School controls				
Peer Socioec. Index	0.162	0.668	-3.259	1.799
Ln School size	5.801	0.925	1.792	7.603
Pupil-Teacher ratio	23.07	6.74	5.13	47
% teachers tenured	0.776	0.347	0	1
Basic resources	4.310	0.808	0	5
Private school	0.203	0.402	0	1
Rural school	0.127	0.333	0	1
Local area controls				
Local density	9290	8780	1	37233
Local Socioec. Index	0.126	0.646	-2.597	1.427
% in Private schools	0.324	0.235	0	1

Notes: School data from years 2009, 2010, 2012 and 2013 and 2016. Includes only Spanish medium primary schools. Socioeconomic index is measured 6 years later from secondary school data. % teachers tenured refers to a permanent versus temporary contract. All local area variables are computed from pupil populations in schools within 2km of the pupil's school (excluding the school itself). Local density is the number of primary school pupils. Local socioeconomic index is computed from the populations of secondary school pupils in schools within 2km of the pupil's school.

Table 3 Learning premium from local density – Reading (2^{nd} grade)

	[1]		[2]		[3]		[4]	
	Reading		Reading		Reading		Reading	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln Local density	10.78**	[54.96]	4.21**	[22.79]	-2.48**	[-10.45]	-2.60**	[-8.63]
Pupil controls								
Socioec. Index			24.51**	[70.42]	13.28**	[87.50]	13.22**	[82.79]
Female			8.41**	[26.23]	7.29**	[27.99]	7.28**	[27.79]
Spanish native language			32.83**	[33.28]	25.26**	[26.22]	25.11**	[25.91]
School controls								
Peer Socioec. Index					35.73**	[41.08]	34.69**	[31.00]
Ln School size					3.09**	[5.22]	3.02**	[5.08]
Pupil-Teacher ratio					0.53**	[8.64]	0.53**	[8.66]
% teachers tenured					14.82**	[11.32]	14.69**	[11.25]
Basic resources					3.16**	[7.88]	3.16**	[7.86]
Private school					7.14**	[4.88]	7.02**	[4.74]
Rural school					6.50**	[6.51]	6.42**	[6.34]
Local area controls								
Local Socioec. Index							-0.6	[-0.40]
% pupils in Private schools							8.83**	[4.33]
Observations	831,296		831,296		831,296		831,296	
R-squared	0.09		0.15		0.18		0.18	

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. Data from years 2009, 2010, 2012 and 2013. All models control for Year dummies and cluster the standard errors at school level.

Table 4 Learning premium from local density – Mathematics (2nd grade)

	[1] Mathematics		[2] Mathematics		[3] Mathematics		[4] Mathematics	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln Local density	8.59**	[30.89]	1.91**	[6.53]	-5.51**	[-14.28]	-4.89**	[-9.94]
Pupil controls								
Socioec. Index			24.85**	[50.17]	14.85**	[74.38]	14.99**	[69.88]
Female			-9.55**	[-18.88]	-10.91**	[-25.95]	-10.90**	[-25.95]
Spanish native language			36.59**	[25.52]	27.91**	[19.24]	28.13**	[19.34]
School controls								
Peer Socioec. Index					39.80**	[28.02]	42.21**	[23.02]
Ln School size					5.35**	[5.28]	5.17**	[5.09]
Pupil-Teacher ratio					0.71**	[6.75]	0.66**	[6.33]
% teachers tenured					23.81**	[10.71]	24.39**	[10.92]
Basic resources					4.79**	[7.14]	4.97**	[7.40]
Private school					-3.64	[-1.44]	-4.76+	[-1.86]
Rural school					15.33**	[9.37]	14.65**	[8.81]
Local area controls								
Local Socioec. Index							-6.83**	[-2.76]
% pupils in Private schools							9.95**	[2.94]
Observations	831,296		831,296		831,296		831,296	
R-squared	0.03		0.07		0.1		0.1	

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. Data from years 2009, 2010, 2012 and 2013. All models control for Year dummies and cluster the standard errors at school level.

Table 5 Summary statistics for value added between primary and secondary schools for rural-urban movers, 2009-2019.

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.

Table 6 Rural-urban migration and value-added in learning between primary and secondary school

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

Table 7 Rural-urban migration and value-added learning, heterogeneity by language and sex, 2009-19.

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

Table 8 Rural-urban Migration Patterns Observed for Current Teenagers in Peruvian Census data

Set	**Birth**	Move?	**Age 11-13**	Move?	**Age 16-18**	Time in urban	Status?	Obs.
1	Rural	No	Rural	Yes	Urban	Short	**Control**	**13785**
2	Rural	Yes	Rural	Yes	Urban	Short	Not used	1286
3	Rural	Yes	Urban	Yes	Urban	Long	Not used	4788
4	Rural	Yes	Urban	No	Urban	Long	**Treat**	**50890**

Notes: Pooled data from Censuses 2007 and 2017. At the age of 16-18, Urban location is defined by current location of residence. Urban/rural status at birth and at the age of 11-13 is defined by district of birth and district of residence 5 years ago. If the district was more than 50% urban in 2007, it is defined as an urban, otherwise rural.

Table 9 Summary statistics for 16-18 year old rural-urban migrants (census data)

	Treatment			**Control**			Difference
	Obs	Mean	S.E.	Obs	Mean	S.E.	z-stat
Studied beyond primary	50,890	0.923	0.267	13,785	0.898	0.302	9.35
Correct age to grade	50,890	0.614	0.487	13,785	0.583	0.493	6.58
Secondary graduate	50,890	0.497	0.500	13,785	0.473	0.499	5.01
Studying beyond secondary	50,890	0.144	0.351	13,785	0.130	0.336	4.16
Age	50,890	17.00	0.82	13,785	16.94	0.81	7.64
Female	50,890	0.487	0.500	13,785	0.493	0.500	-1.20
Mother's age	50,890	42.54	6.70	13,785	42.36	6.77	2.79
Mother beyond primary	50,890	0.361	0.480	13,785	0.346	0.476	3.13
Mother beyond secondary	50,890	0.104	0.305	13,785	0.100	0.300	1.23
Mother speaks Spanish	50,890	0.647	0.478	13,785	0.657	0.475	-2.07

Notes: Sample includes urban teenagers aged 16-18 in censuses of 2007 and 2017, whose families have migrated internally. 'Treatment' refers to people who moved from rural to urban area more than 5 years ago. 'Control' refers to people who made such move at most 5 years ago.

Table 10 Time spent in urban environment and school attainment (census data)

	[1] Beyond Primary	[2] Correct grade to age	[3] Secondary Graduate	[4] Beyond secondary
Treat	.0253**	.0297**	.0137**	.00511+
	[.0028]	[.00452]	[.00434]	[.00308]
Census 2017	.0347**	.101**	.0869**	.012**
	[.00221]	[.00385]	[.00372]	[.00269]
Female	.00511*	.0538**	.0449**	.0427**
	[.00212]	[.00366]	[.00353]	[.00257]
Mother's age	-.000829**	-.00204**	-.00109**	-.000465*
	[.000171]	[.000285]	[.000272]	[.000192]
Mother beyond primary	.0491**	.156**	.137**	.0682**
	[.00224]	[.00441]	[.00432]	[.00331]
Mother beyond secondary	.00979**	.0737**	.0651**	.0934**
	[.00285]	[.00616]	[.00635]	[.00582]
Mother speaks Spanish	0.000882	.0225**	.0227**	.0108**
	[.00272]	[.00517]	[.00503]	[.00369]
Age 17	.012**	-.0435**	.294**	.144**
	[.00264]	[.0046]	[.00436]	[.00242]
Age 18	.0147**	.123**	.459**	.288**
	[.00266]	[.00447]	[.00423]	[.00312]
Constant	.887**	.495**	.148**	-.0524**
	[.00822]	[.0138]	[.0131]	[.00915]
Current district FE	Yes	Yes	Yes	Yes
Observations	64,675	64,675	64,675	64,675
R-squared	0.0774	0.133	0.234	0.17

Notes: '+': p<0.1, '*': p<0.05, '**': p<0.01. Robust standard errors in brackets. Notes: Sample includes urban teenagers aged 16-18 in censuses of 2007 and 2017, whose families have migrated internally. 'Treat' indicates longer exposure to urban environment (more than 5 years) than the reference group, who migrated to urban area at most 5 years ago.

Table 11 7-18 year old urban residents who lived in a different location 5 years ago, censuses 2007 and 2017 (census data)

	Estimation sample: Recent movers to urban area				All urban residents aged 7-18	
	Mean	SD	Min	Max	Mean	SD
Census 2017 (vs 2007)	0.454	0.498	0	1	0.467	0.499
Exposure / Age difference	4.18	2.37	1	11		
Correct grade to age	0.727	0.445	0	1	0.748	0.434
Child in school	0.957	0.202	0	1	0.915	0.279
Origin rurality share (district)	0.226	0.311	0	0.987		
Female	0.493	0.500	0	1	0.494	0.500
Age	10.5	2.6	7	17	12.8	3.4
Mother's age	38.5	6.1	21	61	39.2	7.1
Mother educ secondary	0.660	0.474	0	1	0.716	0.451
Mother educ post-secondary	0.298	0.457	0	1	0.344	0.475
Mother speaks Spanish	0.820	0.384	0	1	0.828	0.377
Observations	175,010				1,413,027	

Table 12 Effect of exposure to urban area (census data)

Dependent: Correct grade to age	[1]		[2]	
	Coef.	S.E.	Coef.	S.E
Exposure × Origin rurality	.0216**	[.00354]	.00993**	[.00144]
Exposure 2	.0466**	[.014]	.0485**	[.00461]
Exposure 3	.0658**	[.0141]	.0563**	[.00481]
Exposure 4	.0672**	[.0157]	.048**	[.00501]
Exposure 5	.0887**	[.0179]	.0507**	[.00521]
Exposure 6	.101**	[.0207]	.0474**	[.00554]
Exposure 7	.125**	[.0241]	.0469**	[.00599]
Exposure 8	.156**	[.0275]	.0452**	[.00657]
Exposure 9	.172**	[.0315]	.038**	[.00732]
Exposure 10	.203**	[.0363]	.0412**	[.00873]
Exposure 11	.188**	[.0434]	.0344**	[.0118]
Female	.016**	[.00614]	.0205**	[.00211]
Family Fixed Effect	Yes			
Origin Fixed Effects			Yes	
Observations	173,662		173,662	
R-squared	.827		.0506	

Notes: '+': p<0.1, '*': p<0.05, '**': p<0.01. Robust standard errors in brackets. All models include age effects. Sample includes urban children aged 7-18 in censuses of 2007 and 2017, whose families have migrated internally within last 5 years.

Table 13 Effect of exposure to urban area - alternative estimations (census data)

	[1]		[2]	
	Correct grade to age		Child in school	
	Coef.	S.E.	Coef.	S.E
Exposure × Origin rurality			.00798**	[.00186]
Exposure 2 × Origin rurality	-.00927	[.038]		
Exposure 3 × Origin rurality	.0169	[.0375]		
Exposure 4 × Origin rurality	.0174	[.0386]		
Exposure 5 × Origin rurality	.0485	[.0392]		
Exposure 6 × Origin rurality	.0571	[.0408]		
Exposure 7 × Origin rurality	.103*	[.0437]		
Exposure 8 × Origin rurality	.13**	[.048]		
Exposure 9 × Origin rurality	.159**	[.0527]		
Exposure 10 × Origin rurality	.232**	[.0649]		
Exposure 11 × Origin rurality	.23*	[.0897]		
Exposure 2	.0542**	[.0168]	.0154*	[.00679]
Exposure 3	.0723**	[.0168]	.0377**	[.00764]
Exposure 4	.0797**	[.0181]	.0658**	[.00943]
Exposure 5	.0985**	[.02]	.093**	[.0117]
Exposure 6	.115**	[.0226]	.118**	[.0141]
Exposure 7	.132**	[.026]	.147**	[.0167]
Exposure 8	.161**	[.0294]	.167**	[.0194]
Exposure 9	.175**	[.0334]	.198**	[.0223]
Exposure 10	.19**	[.0392]	.228**	[.0254]
Exposure 11	.183**	[.0479]	.249**	[.0296]
Female	.016**	[.00614]	.00405	[.00284]
Family Fixed Effect	Yes		Yes	
Observations	173,662		173,662	
R-squared	.827		.83	

Notes: '+': p<0.1, '*': p<0.05, '**': p<0.01. Robust standard errors in brackets. All models include age effects. Sample includes urban children aged 7-18 in censuses of 2007 and 2017, whose families have migrated internally within last 5 years.