DISCUSSION PAPER SERIES

IZA DP No. 18177

The Urban Learning Premium – Evidence from Peru

Sonja Fagernäs
Panu Pelkonen
Juan Manuel del Pozo Segura

OCTOBER 2025

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18177

The Urban Learning Premium – Evidence from Peru

Sonja Fagernäs
University of Sussex

Panu Pelkonen
University of Sussex and IZA

Juan Manuel del Pozo Segura
Pontificia Universidad Catolica del Peru (PUCP)

OCTOBER 2025

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Urban Learning Premium – Evidence from Peru*

Persistent regional inequalities in education and rapid urbanization are common features of emerging economies. We examine the urban learning and schooling premium in Peru using two approaches: 1) estimating the effect of local population density on the value-added in learning with a register of pupils between grades two to eight, and 2) quasi-experimental census-based estimations on the effect of the duration of urban exposure in childhood on school attainment for rural-urban migrants. Unconditional estimates show that a ten-fold increase in urban population density is associated with around 0.13 standard deviations higher value-added in learning. Conditional estimates reveal that this learning premium is explained mainly by the socio-economic status of pupils' households, and local area specific factors, namely the number of nearby schools, size of schools and local wealth levels. School resources, on the other hand, do not matter. This suggests that sorting together with indicators of agglomeration benefits are the key drivers behind the urban density premium in learning. However, once all observable factors are accounted for, the density coefficient turns negative, implying that urban density is also coupled with unobserved factors that are detrimental for learning. The quasi-experimental estimations confirm a positive relationship between urban exposure and educational outcomes; having spent a longer period of education in an urban area leads to higher a likelihood of continuing beyond primary schooling, being enrolled in the correct grade to age, being a secondary school graduate, and studying beyond secondary school.

JEL Classification: I21, O15, R58, H75

Keywords: urbanization,learning, neighborhoods, agglomeration, Latin America

Corresponding author:
Panu Pelkonen
Jubilee Building
University of Sussex
Falmer, BN1 9QU
UK

E-mail: p.o.pelkonen@sussex.ac.uk

* This research was funded by the British Academy's Humanities and Social Sciences Tackling Global Challenges Programme, supported under the UK Government's Global Challenges Research Fund. We thank UMC/MINEDU in Lima for access to the key data. We'd like to thank in particular Rafael González-Val, Santiago Cueto, Alan Sanchez, Manuel Marcos, Sandra Carrillo, seminar participants at GRADE, at the Universidad Antonia Ruiz de Montoya, at the Economics of Education Association Conference in Valencia and our colleagues for helpful feedback. The study has also benefitted from discussions with a number of Peruvian academics and policy makers. The authors take responsibility for any remaining errors.

1 Introduction

Rapid urbanization is a common phenomenon across many low and middle-income countries, together with rural-urban inequality in learning outcomes and the quality of schooling. While there is now an established literature on the urban agglomeration benefits for economic activity and wages (e.g. Glaeser and Gotlieb, 2009), the role of urban density in the production function for learning and skills remains understudied, especially in developing and emerging economy context.

In this study, we analyze the effects of urban density and urban exposure on learning and educational attainment in Peru between 2007-19. We rely on geo-coded, longitudinal register data on schools and pupils, as well as population census data and conduct two separate analyses. The register data allow us to analyze the drivers behind the urban density premium in learning, which have so far received little attention in the related literature.

The benefits of urbanization for economic activity ultimately derive from increased proximity of people, goods and ideas (Glaeser, 1998). Within the context of schooling, urban density tends to be associated with a larger number of schools, teachers and pupils in close proximity. Urbanization and urban density can influence learning via several channels, relating to the quality of schools and the school system itself and to factors beyond the school system. The former will include school specific factors, such as resources, management, and the numbers and quality of teachers. The latter can incorporate pupil level factors such as family socio-economic status and differences in the valuation of education in urban versus rural areas due to differing returns. It will also include urban area or agglomeration factors such as competition, increased scope for co-operation and the spread of teaching and learning technologies, a broader choice over schools and more negative factors such air pollution. The role of policy or policy prescriptions will vary depending on the key channels.

One study that specifically focuses on the effect of urban density on learning is Gibbons and Silva (2008). In the context of England, the authors compare pupils from the same primary school who transition to secondary schools in lower versus higher density areas and find a small, but significant positive effect on test scores, conditioning on other factors. Interestingly, in their context, the unconditional correlation between urban density and test scores is in fact negative. The study highlights the importance of school competition, as it finds that the number of nearby schools has a more robust effect on the test scores than population density. Van Maarseveen (2021) on the other hand focuses on the connection between urban density and the demand for education. He shows that children growing up in more dense areas in the Netherlands were more likely to opt for more academic tracks, controlling for their cognitive ability.

A related strand of literature analyzes neighborhood effects (see e.g. Aaronson, 1998, Potnick et al. 1999, and Chetty and Hendren, 2018). Using tax record data and sibling comparisons in the US context, Chetty and Hendren (2018) find that children who spend more time growing up in a better neighborhood have better outcomes later in life, including college attendance rates.

Outside of high-income countries, the effects of urban density on learning could be speculated to be larger. However, the question is understudied, likely due to a lack of suitable geo-coded data on schools. until recently. Living standards and resourcing of schools can vary greatly between urban and rural areas, as can the demand for education. Furthermore, in terms of the labor markets of teachers, the literature for developed countries often finds that teachers don't prefer dense cities as places of employment (see e.g. Chevalier et al 2007), whereas in the developing and emerging economies, the opposite has been found; urban school locations are strongly favored by teachers, especially by female teachers (Fagernäs and Pelkonen 2012, Evans and Acosta 2023).

One study focusing on developing countries is Van Maarseveen (2025), who relies on population census data in African countries. He finds that childhood urban exposure raises primary school completion and literacy rates. As potential channels he suggests higher returns to education in urban areas, lower travel costs, higher opportunity costs in urban areas and income effects. 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 is associated with better learning outcomes, especially for under 8-year olds.

Our study provides several contributions to the small, existing literature on the topic. Firstly, we focus on a middle-income country and establish that an urban learning premium exists, using two different methods. Secondly, we analyze a range of drivers behind the premium. A number of studies show that migrants to urban areas are more educated or have higher cognitive ability (e.g. Bacolod et al. 2010, De la Roca 2017, Bütikofer and Peri, 2021). Given the relatively rich data on schools, families and the local environment, we are able to distinguish between school specific factors, the sorting of pupils and families versus more place-based', or urban agglomeration factors. The few existing studies remain quite speculative on the factors that lie behind an urban learning premium, possibly due to the nature of the data used.

We focus on primary and secondary school aged children, and conduct two separate analyses. In the first analysis, instead of a simple urban-rural divide and retrospective data, we study the association between a refined measure of urban density and current learning. Based on pupil numbers, we create a measure of urban density within a 2km radius of each school ('micro-locality') and estimate its association with pupil specific test score gains between primary and secondary school (grades 2 to 8) for four cohorts of pupils between 2009-19. The data allow us to analyze the relevance of a number of potential drivers behind the urban density premium in learning.

The unconditional estimates suggest that on average, a one log-point increase in urban density is associated with a 0.06 standard deviation higher value-added in learning between grade 2 and 8. A conditional analysis indicates that both local area specific factors that reflect urban 'agglomeration' benefits, as well as sorting, mainly household socio-economic status, are the key drivers behind the urban density premium in learning. The local area specific factors included are local wealth, the number of nearby schools and school size. On the other hand, key school characteristics, such as the pupil-teacher ratio and school resources, are largely unimportant in explaining the density premium. Once all these factors are controlled for, the coefficient on urban density turns negative, implying that density is also associated with unobserved characteristics that are detrimental for learning, with air pollution being one plausible example.

In the second analysis, we rely on a quasi-experimental approach and use population census data to compare educational outcomes of teenagers who have spent different proportions of their lives in an urban area. The approach is similar to the one used by Chetty and Hendren (2018) and Maarseveen (2025). In the Appendix, we also include an analysis based on sibling comparisons. Our results indicate that a longer exposure to an urban location leads to a nearly 3 percentage points higher likelihood of studying beyond primary school. It also leads to a 3 percentage points higher likelihood of attending the correct grade for age, and a 1 percentage point higher likelihood of graduating from secondary school. While this analysis cannot inform us of the channels behind the urban premium, it confirms the existence of a significant urban educational premium in Peru, using a plausible control group. It also provides an indication of the implications of urban areas for longer term educational outcomes.

The paper is organized as follows. Section 2 describes the data. Section 3 focuses on the association between urban density and learning, including an analysis of the potential channels. Section 4 presents the analysis on the impact of urban exposure on broader schooling outcomes, based on a Census sample of migrants. Section 5 concludes.

2 Data and general descriptive statistics

For data on schools, we use the Censo Escolar, which is an annual school level register data set, which covers all schools in Peru. Among other things, these school data include school level aggregates on resources, teachers, pupils and location, but do 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 test scores for second and eighth grade pupils from the Evaluación Censal de Estudiantes (ECE). This includes nationally comparable data for Reading and Mathematics, and some key characteristics on pupils and their background. We restrict the analysis to Spanish speaking schools, as bi-lingual schools, which are largely rural, tested pupils at a different age.

More specifically, we rely on a large separate ECE sub-sample, which tracks pupils in the school census from grade 2 to 8 for four cohorts: those who are in grade 2 in years 2009, 2010, 2012 and 2013, and in grade 8 in 2015, 2016, 2018 and 2019. While the dataset includes all pupils in grade 8, depending on cohort, around 75-80% of the pupils' primary schools are traced, which leads the panel sample to be somewhat biased to urban areas. The key advantage of using this sub-sample instead of the full ECE data set is that the panel dimension allows us to account for previous learning levels, and as such, allows us to de facto control for the initial level of skills.

Additionally, we use the national Census, Censo de Población y Vivienda, for 2007 and 2017, which is representative at the district level. Peru has 25 regions and around 1,800 districts. We use the census for the analysis on the impact of urban exposure on school progression and attainment of children and teenagers in Section 5. The Census data are available for the entire population.

We measure the local population density as the natural logarithm of the number of primary school pupils that are enrolled in primary schools within a 2km radius of the pupil's own primary school. The pupils are currently in grade 8, but our density indicator is based on the coordinates of the primary rather than the secondary school, because the pupils will have spent a longer time in the primary school and are often likely to reside closer to the primary school. In the econometric analysis below, we also include a separate indicator for the number of schools within a 2km radius of the primary school, primarily to reflect competitive pressure faced by schools, and distinguish this from the possible density effect.

Figure 1 displays a simple relationship between the average Reading scores for grade 8 pupils against the indicator of local pupil density for 2016. For this descriptive purpose, the full census of secondary schools is used (8^{th} grade scores are available for pupils in all schools). The left-hand side image uses a linear scale for the x-axis, while the right-hand side uses the natural logarithm of the number of pupils within 2km. The density measure varies quite dramatically across schools and is strongly correlated with learning outcomes. The left-hand side image shows that average learning outcomes increase rapidly with urban density until the number of pupils reaches around 15000 with a 2km radius, after which the relationship flattens. This would correspond to a dense urban area. The pattern indicates that there is a clear urban density learning premium. The right-hand image shows that the relationship is nearly log-linear throughout, apart from the very highest densities. Interestingly, this is in stark contrast with Gibbons and Silva (2008), who find a negative unconditional correlation for England.

More generally, Figure 2 indicates that second grade test scores have improved since 2007 in both urban and rural areas, but despite convergence of school resources, rural areas have not caught up with urban areas. In fact, the gap appears to have slightly widened with respect to Mathematics. Here districts are defined as 'urban' if their rate of urbanization was over 80% in the Census of 2007 (391 districts), and 'rural', if below 80% (1373 districts).

3 The effect of urban density on learning

3.1 Conceptual framework

We consider a production function for skills, where learning is a product of school characteristics, family inputs, and the surrounding environment. Adapting Glewwe and Muralidharan (2016), we conceptualize the learning production function as follows

(1) $L = f(S, C, H, I, U)$,

where L is learning, S is a vector of school and teacher characteristics, C is a vector of child characteristics, H is a vector of household characteristics, and I is a vector of school inputs under the control of households, such as children's daily attendance/time use, effort in school and in doing homework, and purchases of school supplies. We then enhance the standard framework by adding the component 'U' to equation (1), which represents local area specific factors connected with urban density, or 'agglomeration benefits' not captured by the other components.

We consider the relevant local area to be the one corresponding to the 'lived experience' of pupils, encompassing the area around the pupil's school and home. U incorporates factors such a higher degree of competition between schools and increased scope for co-operation and transmission of teaching and learning practices between schools, due to a higher number of schools in close proximity. It also includes the socio-economic status of the local area, which reflects the wealth of the area, but also the broader peer group of the pupil. While school size is in principle a school characteristic (S), larger schools may only be enabled by a denser urban environment, reflecting a further potential agglomeration benefit.[1] Larger schools allow for more specialized staff, and more varied facilities, which may not be captured by standard measures of school inputs, such as pupil-teacher ratios. Larger schools may also be able to direct a larger share of employees to teaching as opposed to the management of the school. Finally, U can also capture the existence of a denser market for teachers and potential negative features of urban areas, such as air pollution.

In addition to the factors captured by U, an urban environment can be associated with differences in all the other components as well. For example, schools may be better equipped, children may

[1] There is a clear positive correlation between our density indicator and school size. The correlation between Ln(Density) and the log of primary and secondary school size is 0.52 and 0.48 in our sample, respectively.

be healthier, households wealthier and more educated and general valuation placed on education may differ. A harder to capture feature in the context of developing countries is that families' demand for education may increase as they move to cities. Returns to education are likely to be higher in more urban contexts (eg. Backman, 2014, Gao and Li 2022) and parents and children will understand that education is the path to increased opportunities to modern occupations and higher incomes. This implies that even when conditioning for parental education and resources, urban areas may make children exert more effort on mastering skills. This effect may be partially visible in measures such as children's time use for schoolwork and hiring of private tutors, for which we do not have indicators in the data sets.

Education policy can only directly address some of the factors behind a possible urban learning premium. In particular, school specific characteristics are relatively easy to change, while the selection of families and area-specific factors are likely to be beyond the control of education policy.

3.2 Estimation: urban density and learning

To study the association between urban density and learning, we rely on a pupil-level lagged score value-added model.[2] We begin by estimating the unconditional association between the value-added in learning between grades 2 and 8 and the local population density, measured by the number of pupils in primary schools within a 2km radius of the pupil's primary school. The estimated coefficient will be a combination of all the components in equation (1), to the extent that they correlate with our measure of urban density. We then add sets of explanatory variables in stages to investigate how the association between learning and urban density changes once specific components are controlled for. The coefficient on urban density would be expected to fall as observable pupil and school characteristics are added, and potentially be reduced to zero if the observable variables (on agglomeration benefits) have strong predictive power. This approach allows us to gauge the drivers behind the observed urban density premium, based on variables that are available to us over time.

We do not know the location of the pupils' home, just that of the school. We therefore restrict the sample to pupils whose primary and secondary schools are no more than 2km apart, as for this sample it is reasonable to assume that the pupils haven't changed their area of residence. Therefore, the density based on the location of the primary school is the relevant 'treatment' density for the pupil throughout the examined period. For most pupils in the dataset (79.5%), the

---

[2] For properties and some benefits of such models in the context of program evaluation, see eg. Andrabi et al (2011).

primary and secondary schools are within 2 kilometers of each other.[3] In addition, had the family moved, we would not know the point at which this happened between grades 2 and 8, and therefore whether the density around the primary or the secondary school is the more 'relevant treatment' for the pupil in question. A majority of the value-added between grades 2 and 8 will have been accumulated over the primary years 2-6. While this sample selection constrains generalizability somewhat, it guarantees that the pupils have spent the period of interest essentially in the same neighborhood, subjected to the same population density.

We estimate the following model for both normalized Reading and Mathematics scores:

$$(2)\quad Score8_{iscl,} = \alpha Score2_{ipcl,} + \beta_1 \ln\left(Density_{pc}\right) + \boldsymbol{C_{ic}\lambda} + \boldsymbol{Z_{\{p,s\}c}\theta} + \boldsymbol{X_l\beta} + \delta_c + \varepsilon_i$$

where *i* refers to individuals (pupils), *s* to secondary and *p* to primary school, *c* to cohorts and *l* to local area. The scores are explained with urban density and a vector of pupil characteristics ($\mathbf{C_{ic}\lambda}$), a vector of school characteristics separately for primary and secondary schools ($\mathbf{Z_{\{p,s\}c}\theta}$) and fixed local area characteristics ($\mathbf{X_l\beta}$). The model controls for cohort effects (δ_c) for the four cohorts that are enrolled in grade 8 in 2015, 2016, 2018 and 2019.[4] All estimated standard errors are corrected for spatial clustering within 2km of primary school locations, using Arbitrary Correlation Regression by Colella et al (2023).

The key coefficient of interest, β_1 measures the density premium in learning value-added "given the other controls". The interpretation of the coefficient thus will depend on which local characteristics and school characteristics are included in the model. We begin with a value-added specification, which in addition to urban density, controls only for the pupil's past test score (grade 2) and cohort effects. In this estimation β_1 will represent the combined impact of the multiple potential sources of the urban premium discussed in the conceptual framework in 3.1, such as home inputs, school inputs and local agglomeration factors. We then add sets of explanatory variables in stages. It is worth noting that when we control for urban density together with components other than those in U in equation (1), β_1 will provide us with an estimate for the 'net' urban agglomeration benefits for learning. This will be approximate as the variables available to us may not capture all possible components accurately; for instance, the data do not

[3] There is a sample of 946.268 secondary school pupils for whom we have primary school information and data on learning outcomes. Out of these, 814.131 pupils have non-missing observations for all key explanatory variables. In this sample the median distance between primary and secondary schools is 0.36 km, and the average is 3.49 km. 646.916 pupils, or 79.5% have at most 2 km between primary and secondary schools.

[4] An alternative to our approach would be to use the full sample and explain changes in value added with changes in population density between the two observations, as in Gibbons and Silva (2008). However, this would be less sensible in our case, as data on movers is likely to be more selected due to imperfect coverage. Furthermore, Gibbons and Silva compare end-of primary school scores to end-of secondary school scores, so that the value added they examine is obtained in the secondary school only.

contain variables associated with component I (school inputs under the control of households), nor potential features of urbanization that may harm learning.

The common correlation between the different factors of interest suggests that a strict 'causal' interpretation of the estimated effects is not warranted. That being said, the value-added model mitigates this concern as Score2 (test scores in grade 2) controls for baseline skills, and therefore for selection arising from any inputs that enter the production function for test scores in grade 2. The main threats to identification therefore arise from unobserved factors that are correlated with additional learning taking place between grades 2 and 8.

In terms of school characteristics, we include indicators for school resources in both primary and secondary schools for which we can obtain comparable data for primary and secondary schools across time; pupil-teacher ratios (PTR), whether the school is private or public, the presence of an internet connection, availability of (any) computers and the number of pupils in the school (school size). In the absence of comparable data on teacher qualifications over time, we include the share of primary and secondary teachers with a permanent versus a temporary contract (tenure) as a potential indicator of teacher quality. Most of these indicators can in principle be influenced directly by education policy or funding. School size however, can be considered an indicator of agglomeration benefits, as larger schools become possible only with denser populations.

The data source contains fairly limited information on the pupils' family background. The key household characteristic is the index of socio-economic status in the eighth grade (from ECE, created by the ministry of Education), which we include to reflect potential selection effects. This indicator is not available for the second grade. We also include pupil's gender and whether the pupil's home language is Spanish, although this is the case for the great majority in the sample used (95%).

The vector for local area characteristics, capturing potential agglomeration benefits, includes two variables. The first one is a relative Wealth Index, which is an open-source geospatial indicator at 2.4km resolution published by the Meta corporation. It is not longitudinal, but based on a machine learning algorithm that is trained on household asset wealth in Peruvian Demographic and Health Survey of 2009 using infrastructure data sources such as telecommunications data, night lights and road infrastructure. A high-resolution local wealth levels are then predicted based on the same fine geospatial infrastructure data.[5] We link this indicator to the location of the pupils' primary schools. As a robustness check for the local socio-economic level, we computed the

[5] Source: Chi et al (2022). Link: https://dataforgood.facebook.com/dfg/tools/relative-wealth-index. The data is not longitudinal but is representative of wealth levels in 2010s.

average pupil-level socio-economic index in grade 8 within the 2km radius of a pupil's primary school based on the ECE data, excluding the school itself. We found these two measures of local wealth to be highly correlated (r = 0.82), and leading to very similar results.[6] Secondly, we compute the logarithm of the number of primary schools (both public and private) within the 2km radius of the primary school that the pupil attended. While the first proxies for general wealth and the broader peer group of pupils in the area, the latter can reflect factors such as competition, and the scope for transfer of practices and learning technologies and collaboration between schools in the local area.

Table 1 summarizes the variables used in the estimation. The learning outcomes for Reading and Mathematics are standardized. The mean ln(Density) is 8.0, implying that about 3084 primary aged pupils within the primary schools are located within 2 kilometers of the pupil's secondary school. The standardized home socio-economic status index is 0.137, suggesting a slight sample bias towards wealthier families. This is mostly due to about 20-25% of the secondary pupils not being traced to their primary schools. Pupil-teacher ratios (PTR) in secondary schools are on average 16, while in primary schools they are 24. 16 and 19 percent of pupils attended a private primary or secondary school, respectively. There is significant variation in the number of primary schools within the 2km radius, with an average of 42.6 schools. Urban areas are especially characterized by the co-existence of larger public schools and private schools of varying size.

3.3 Results

Table 2 presents the results on the association between local density and learning outcomes. Panel A relates to Reading and Panel B to Mathematics. Table 2 provides condensed results, and the full results are available in Appendix Tables A1 and A2.

The first column of Table 2 presents the results of a model where the eighth-grade test scores are regressed only on the local density indicator, cohort dummies, and the grade 2 test scores. The connection is strong; a one log point, or a 2.6-fold increase in population density is associated with 0.062 standard deviations higher Reading scores and 0.055 standard deviations higher Mathematics scores. The R-squared shows that 38.1% of the variation in Reading is explained by the model while the corresponding share for Mathematics is 33%. A ten-fold increase in population density would increase Reading scores by 0.062*[2.303] = 0.143 and Mathematics score correspondingly by 0.127 standard deviations. As such, urban density appears to be a strong predictor of learning value-added.

The purpose of the next four columns is to unpack the sources of this urban advantage. In column 2, we add the index for the household socio-economic status, as well as gender and home

[6] The alternative results are available on request.

language. The results indicate that a significant share of the urban density premium is due to selection: while remaining statistically significant, the coefficient for urban density becomes three to four times smaller for Reading and Mathematics.

The inclusion of the school input variables in column 3 (PTR, share of tenured teachers, private school status, internet and computer access, separately for primary and secondary schools that the pupils attended) has practically no impact on the density coefficient. This, while surprising, is not entirely out of line with the literature on the impact of school inputs on learning (see eg. Glewwe and Muralidharan, 2016). This suggests that a policy of equalizing school resources across rural and urban areas could not be expected to lead to a sizeable reduction in urban-rural learning differences.

The coefficient for urban density in the third column is our preferred estimate for the 'net agglomeration benefits' of urbanization on learning. It represents the residual effect once pupil selection and malleable school resources have been accounted for and can reflect both positive and negative features of the urban learning environment**.** Now, a one log point, or a 2.6-fold increase in population density is associated with 0.018 standard deviations higher Reading and 0.013 standard deviations higher Mathematics scores. A ten-fold increase in density would now increase Reading scores by 0.018*[2.303] = 0.041 standard deviations, and Mathematics scores by 0.03 standard deviations. These effects, while significant, are economically less meaningful compared to the urban learning gradient in the first column of Table 2.

In column 4 of Table 2, in addition to controlling for the past test score, home socio-economic status, pupil characteristics and school inputs, we add two local areas characteristics that capture the school environment: the size of the pupil's school, for both primary and secondary schools and the log of the number of primary schools within a 2km radius of the primary school that the pupil attended. We have labeled these variables as proxies for agglomeration benefits, together with local wealth, which is added in the next stage. Interestingly, at this stage, the coefficient on urban density turns significantly negative. The change in the magnitude of the coefficient on urban density suggests that these two proxies of agglomeration capture substantial positive benefits of agglomeration. For Reading, the estimate for urban density changes by (0.018-(-0.025)) = 0.043, which is twice as large as the 'net' benefit of 0.018 in column 3, and similar in size to the change brought about by the inclusion of household socio-economic factors in column 2.

Finally in column 5, adding the indicator for the local wealth to the model leads to an even more negative coefficient for urban density. The local level of wealth and development, as captured by the indicator, is also correlated with faster learning. The full results in the Appendix show that in the final estimation, all indicators of agglomeration remain independently statistically significant for both Reading and Mathematics, apart from the variable indicating the size of the primary school.

The fact that the estimated effect of density turns negative in the final two estimations suggests that urban density or city centers also encompass aspects that are unconducive to learning. Existing literature supports potentially negative effects from factors such as air pollution, for which we do not have geographically precise enough data (on pollution and learning in particular, see e.g. Pham and Roach 2023, Persico and Venator, 2021).[7] Taken together, after controlling for individual and school-specific factors, we find that the net effect of urban density on learning value-added is fairly small, but positive. This is captured by the coefficient 0.018 in column 3. This effect appears to be a sum of two opposing factors; the positive agglomeration benefits for learning which we can proxy with a few variables, and a negative effect, largely related to agglomeration, for which we do not have data. The result indicates that the positive effects outweigh the negative agglomeration effects.

On the other hand, differences in school resources do not appear to be a key driver of the urban density premium in learning, despite the common perception that urban areas have better school resources and infrastructure. Larger school size does predict learning positively, but larger schools are typical of more densely urbanized areas.

From a policy perspective, it is worth noting that the key drivers cannot be manipulated easily by policy in a less dense or rural context, even with generous funding. It may simply not be feasible to have large schools in the countryside, nor is it possible to have intense school competition, as in urban areas. The local level of wealth and infrastructure could in principle be more readily influenced by economic redistribution.

4 Urban-rural migration and educational gains – Census estimates

4.1 The empirical approach and data

As an alternative analysis, which also allows us to study slightly longer-term effects, we study the effects of an urban environment on several educational attainment outcomes using a sample of internal migrants. Here we can only focus on the urban-rural distinction and not the density gradient.

Rural-urban migrants are a selected group of people, both along observable and unobservable dimensions. To mitigate the role of this selection on the estimates, we rely on a similar approach as used by Chetty and Hendren (2018) to study the causal effect of a neighborhood on later outcomes in the US. Van Maarseveen (2025) also relies on a similar idea in a development context

[7] Sufficiently dense measurements for pollution for the whole of Peru do not exist for our purposes.

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 migrants 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 use a similar identification approach, with additional features such as a specification that focuses on a sibling comparison. Such an approach can be argued to be more rigorous since it allows for the inclusion of family fixed effects.

We utilize these ideas with data for the 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 restrict the analysis to rural-urban migrants and therefore focus on children or youth who have moved from rural to urban areas. In our main analysis, we compare the educational outcomes of 16-18 year olds who moved to urban areas more recently to the outcomes for those who have spent a longer time in urban areas. In the Appendix, we further report the results for sibling comparisons, but this does not form our main analysis, as the age difference between siblings implies that we can only use one educational outcome indicator. The conclusions are nevertheless similar to those of the main analysis.

The Peruvian Census allows us to identify children in internal migrant families by comparing their district of birth to the current one. We classify districts into urban and rural, based on a threshold of 50% of the population being urban versus rural. The district of birth is defined based on the district in which the child's mother lived 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 sort internal migrants by the length of time they have spent in the current urban location.

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 the last 5 years, the family is classified as a 'Recent migrant' 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.

We compare teenagers who have moved to urban areas across households. We focus on ages 16-18 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.[8] In our sample of internal migrants, 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 3 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 definition of the "treatment" category is based on the length of residence in an urban location. Those who moved more recently to an urban area are in the "Control" group. In the analysis, we thus compare sets 1 and 4 in Table 3. Sets 2 and 3 which have been excluded from the analysis have moved at least twice and have small samples.

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

$$(2) EduOutcome_{icad} = \alpha + \beta Treat_{ic} + \boldsymbol{X}_{ic}\theta + \pi_c + \sigma_a + \gamma_d + \varepsilon_{ic},$$

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. While the two groups are observationally quite similar, leading to a quasi-experimental interpretation, we still control for age, language and education of mothers (vector $\mathbf{X}_{ic}$). 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.

The interpretation of β is based on the distinction that we only know whether the teenager has spent "more or less than 5 years" in the current urban location. Had the children moved at the midpoint of the possible timing windows, the control group would have spent on average 2.5 years in an urban area and the treatment group 11 years. As such, the estimated effects are due to roughly 8.5 more years in an urban area, which for these children would coincide largely with the time they spend in primary school.

A key potential concern in the census analysis above is that the selection of families that moved later might be different to those who moved earlier. As a robustness check, in the Appendix, we provide an additional estimation where we compare siblings from the same families, who have different exposure time to urban environment. The main drawback of the approach in our case is

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

that as the age range of the siblings is extended to range from 7 to 18, we have only two comparable outcome variables; whether the pupil is in the correct grade for age and whether the child attends school.

4.2 Results

Table 4 provides summary statistics for our sample of urban internal migrants, pooled from the Censuses of 2007 and 2017. The last column refers to the 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-randomized setting. However, there is a significant difference with respect to whether the mother speaks Spanish and whether she has studied beyond primary education between the two groups. For the outcome variables, the statistical differences between treatment and control are much stronger. The final column of Table 4 also shows the means for the non-migrant families who have stayed put in districts defined as being more than 50% rural. The differences compared to migrants are stark and highlight the nature of selective migration. Only 15.2% of non-migrant mothers have studied beyond primary education, and only 2.5% beyond secondary education. For migrants (treatment and control), the equivalent numbers are around 35% and 10%, respectively. Migrants are also much more likely to be Spanish speakers.

The results are presented in Table 5. A longer exposure to an urban location leads to an approximately 2.5 percentage points higher likelihood of studying beyond primary. It also leads to a 3 percentage points higher likelihood of attending the correct grade for age, and a percentage point higher likelihood of graduating from secondary school. There is a small, marginally significant effect on being enrolled in studying beyond secondary school.

These results are consistent with the ones presented in Section 3. In the previous analysis we found that a significant urban rural premium exists, which is partly explained by urban location specific characteristics, such the differences in average school size, competition and potential transfer of practices between schools, as measured by the number of primary schools within 2km per pupil population. The quasi-experimental results in the current section provide further evidence that such urban premium is likely to be causal.

4.3 Robustness check using siblings

The Appendix contains an analysis of siblings who have spent a different proportion of their schooling in an urban area. The sample is smaller, and the number of comparable outcome variables is more limited. The average age difference between siblings is 4.18 years, and we conclude that the younger siblings, who spent a larger part of their education in urban areas, are more likely to be in the correct grade for age. They are also more likely to attend school (controlling for age). These results, which account fully for family fixed characteristics, are in line with the results of the main census analysis.

5 Conclusion

Across the world, there is relatively little systematic documentation on the impact of the urban environment on learning and schooling, and the fundamental reasons behind it. This study provides a systematic analysis of the differences in learning and educational outcomes of children across the rural-urban dimension in Peru, and of the factors that explain why more dense areas are associated with more learning. We employ the Peruvian school census, household surveys and censuses from the period of 2007-2019, and conduct two complementary pieces of analysis.

Firstly, we examine the relationship between geo-coded urban density and value-added test scores of school pupils, both unconditionally and including pupil, school and local area controls. The analysis uses longitudinal data for four cohorts of pupils tested in grades 2 and 8. As such, the identification of the effects is aided by value added models, which allow us to account for pupil-specific factors with past test scores.

The first key finding is that urban environment speeds up learning substantially, as measured by value-added scores; a ten-fold increase in population density is associated with a 0.13-0.14 standard deviation improvement in learning value-added between grades 2 and 8. We then conduct a set of conditional estimations, by adding different characteristics in steps. The results indicate that both urban 'agglomeration' benefits, as well as sorting, in particular household socio-economic status, are the key drivers behind the urban density premium in learning. Importantly, school-specific resources such as the numbers of teachers, their status, availability of computers and the internet, or private versus public schools appear to contribute very little to the urban learning premium.

With respect to local area factors, we also find evidence of the presence of negative agglomeration effects on learning. This could reflect factors such as air pollution, for which we do not have precise enough data. Once all measurable factors are accounted for, the effect of population density on the value-added in learning turns negative. This suggests that urban density is not only coupled with benefits, but also unobservable costs, as has been documented in the more established literature on agglomeration effects of cities in labor economics (e.g. Sveikausas 1975, Combes et al 2012). We can therefore assume that the net agglomeration effect consists of two opposing effects; firstly, sizeable positive agglomeration benefits that are proxied by larger school size, the number of schools nearby and local wealth, and secondly, features of urban density that are harmful for learning. In our study, these benefits outweigh the disadvantages of agglomeration.

Further on the agglomeration benefits, larger schools appear to have beneficial effects on learning, which are not captured by available items on teachers and school resources in the data. These can relate to school management, the internal use of resources or more general scale benefits. Similarly, the relevance of the number of nearby schools may reflect several factors, such as school competition or the spread of learning practices and technology. Local wealth can also proxy for the broader peer group of the pupil.

The second complementary analysis in the paper uses census data of 2007 and 2017 to compare teenagers (aged 16-18) who are internal migrants, but have lived in an urban location either for a longer (over 5 years) or a shorter duration (under 5 years). The identification of the effects of urbanization relies on the idea that even if internal migrants are a selected group, comparing internal migrants who moved at different times alleviates the selection bias in the analysis. The children of interest are still living at home and the data allow us to examine four outcomes: the likelihood of continuing beyond primary schooling, being enrolled in the correct grade for age, being a secondary school graduate, and studying beyond secondary school. We find that a longer exposure to an urban environment affects all of these outcomes positively with at least marginal statistical significance. One of the outcomes, correct grade for age, can also be tested using sibling comparisons, and thus accounting for family fixed effects. The results, presented in the Appendix, support the findings.

Overall, the findings from the census-based analysis are consistent with the first analysis – the net effect of an urban environment on educational outcomes is real and statistically significant. Given that the census data reveal that rural-urban migrants are highly selected, both analyses indicate that there is a remaining urban learning premium that cannot be explained simply by the sorting of families.

The results pose a dilemma for policy aiming to reduce urban-rural disparities. Education policy can most directly target school-specific resources, but these do not appear to be the drivers of the urban learning premium, despite urban schools often having better resources. Instead, the benefits of urbanization are attributed to factors that cannot as easily be replicated in areas with a lower population density, namely larger school size, local wealth and a larger number of nearby schools, as well as household socio-economic status.

The urban learning premium and its determinants have only been studied in detail for very few contexts, and even less for emerging and developing economies where the magnitude can be expected to be larger. Further case studies that take the local circumstances into consideration would be warranted. As data sources improve, the role of school management and specialization, as well as school competition and spill-overs in the learning process need to be understood better.

This would ultimately help gauge the educational benefits of urbanization and facilitate the design of educational policies for disadvantaged areas.

6 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). Intergenerational Mobility in Africa. *Econometrica*, 89(1).

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja and & Tristan Zajonc (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, *3*(3), 29-54.

Backman, Mikaela (2014). Returns to education across the urban-rural hierarchy. *The Review of Regional Studies*, *44*(1), 33-59.

Bacolod, Marigee, Bernando S. Blum, and William C. Strange (2010). Elements of skill: traits, intelligences, education, and agglomeration. *Journal of Regional Science*, *50*(1), 245-280.

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.

Bütikofer, Aline, and Giovanni Peri (2021). How cognitive ability and personality traits affect geographic mobility. *Journal of Labor Economics*, *39*(2), 559-595.

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.

Chevalier, Arnaud, Peter Dolton and Steven McIntosh (2007). Recruiting and Retaining Teachers in the UK: An Analysis of Graduate Occupation Choice from the 1960s to the 1990s. *Economica*, *74*(293), 69-96.

Colella, Fabrizio, Rafael Lalive, Seyhun O. Sakalli and Mathias Thoenig (2023). acreg: Arbitrary correlation regression. *The Stata Journal*, *23*(1), 119-147.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, *80*(6), 2543-2594.

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.

Gao, Xing & Li, Maishou (2022). Differences between returns to education in Urban and rural China and its evolution from 1989 to 2019. *Plos one*, *17*(10), e0274506.

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

Glaeser, Edward L. (1998). Are cities dying? *Journal of Economic Perspectives*, *12*(2), 139-160.

Chi, Guanghua, Han Fang, Sourav Chatterjee and Joshua E. Blumenstock (2022). Microestimates of wealth for all low- and middle-income countries. Proceedings of the National Academy of Sciences, Jan 2022, 119 (3) e2113658119; DOI: 10.1073/pnas.2113658119

De la Roca, Jorge (2017). Selection in initial and return migration: Evidence from moves across Spanish cities. *Journal of Urban Economics*, *100*, 33-53.

Evans, David K., and Amina M. Acosta (2023). How to recruit teachers for hard-to-staff schools: A systematic review of evidence from low-and middle-income countries. *Economics of Education Review*, *95*, 102430.

Fagernäs, Sonja, and Panu Pelkonen (2012). Preferences and skills of Indian public sector teachers. *IZA Journal of Labor & Development*, *1*(1), 3.

Glaeser, Edward and Joshua Gottlieb (2009). The wealth of cities: agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, 47(4), 983-1028.

Glewwe, Paul, & Muralidharan, Karthik (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education* (Vol. 5, pp. 653-743). Elsevier.

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

van Maarseveen, Raoul (2025). The effect of urban migration on educational attainment: evidence from Africa. *Available at SSRN 3836097*.

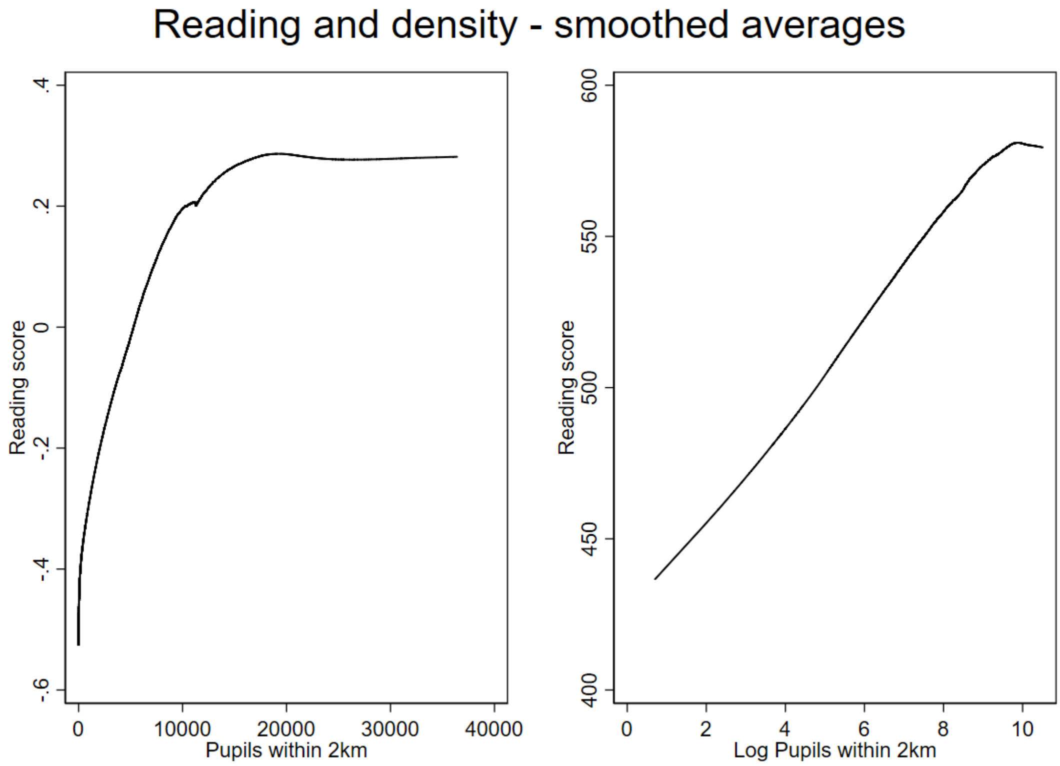
Persico, Claudia L. and Joanna Venator (2021). The effects of local industrial pollution on students and schools. *Journal of Human Resources*, *56*(2), 406-445.

Pham, Linh and Travis Roach (2023). Particulate pollution and learning. *Economics of Education Review*, *92*, 102344.

Sveikauskas, Leo (1975). The productivity of cities. *The Quarterly Journal of Economics*, *89*(3), 393-413.

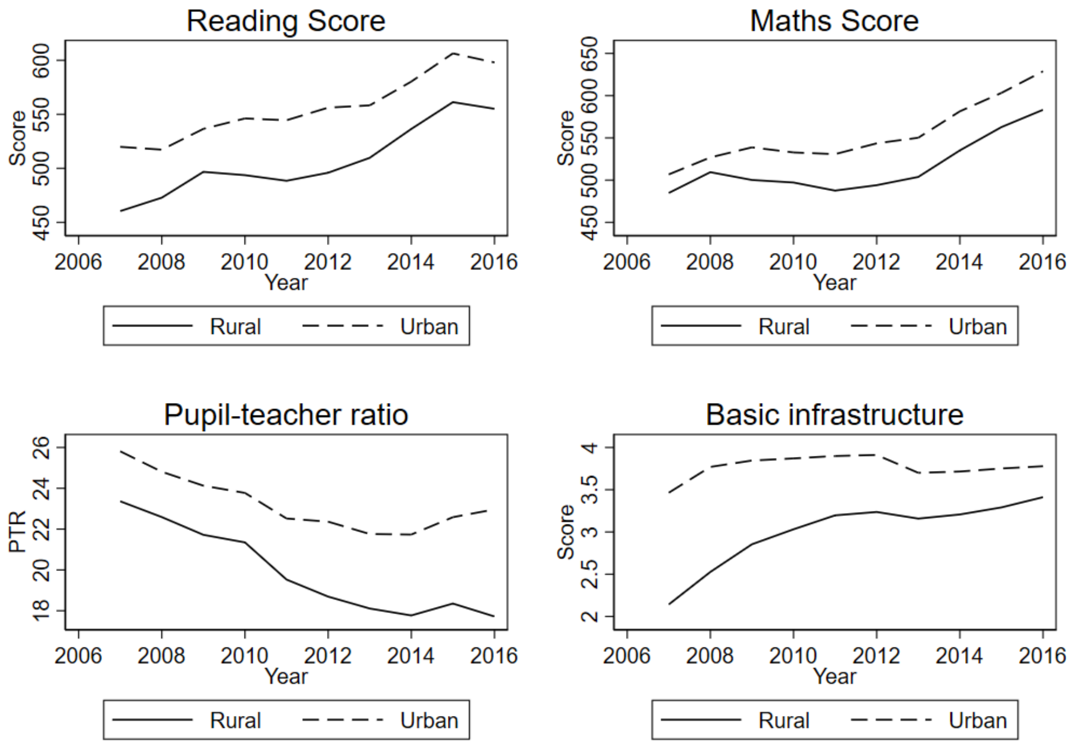
Figures and Tables

Figure 1 Relationship between 8th grade reading scores and pupil density within 2km of the school



Notes: Based on full school census 2016 for secondary schools and 8th 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. Density is based on the number of primary school pupils in all primary schools within 2km of the secondary school.

Figure 2 Primary school second grade test scores and key school resources over 2007-2016 by urban/rural districts.



Notes: Districts are defined as 'urban' if their rate of urbanization 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.

Table 1 Summary statistics for panel of pupils in grades 2 and 8

N = 646914	Mean	S.D.	Min	Max
Reading (z, Grade 8)	0.017	0.989	-6.266	5.149
Reading (z, Grade 2)	0.053	0.988	-5.402	3.046
Maths (z, Grade 8)	0.015	0.992	-5.280	4.848
Maths (z, Grade 2)	0.084	1.008	-4.665	3.666
Female	0.516	0.500	0	1
Spanish speaker	0.955	0.207	0	1
Home SES Index (z)	0.137	0.904	-3.480	9.436
Ln(Density)	8.034	2.041	0	10.513
Local wealth (z)	0.123	0.891	-3.664	1.783
PTR primary	23.78	8.62	0.75	448
PTR secondary	15.70	6.09	0.167	190
% teachers tenured primary	0.863	0.225	0	1
% teachers tenured secondary	0.484	0.319	0	1
Private primary school	0.160	0.367	0	1
Private secondary school	0.192	0.394	0	1
Internet primary	0.545	0.498	0	1
Internet secondary	0.886	0.317	0	1
Computers primary	0.849	0.359	0	1
Computers secondary	0.909	0.288	0	1
Ln(School size) primary	5.931	0.869	0	7.603
Ln(School size) secondary	6.019	0.940	0	8.082
Primary schools within 2km	42.65	41.76	1	191

Notes: The panel data includes only Spanish medium schools and is not a random sample of Peruvian schools. Data is shown for pupils whose location and all variables are non-missing for both primary and secondary school, and who attended secondary school at most 2km away from their primary school. The pupils were in second grade in 2009, 2010, 2012 and 2013, and in 8th grade in 2015, 2016, 2018 and 2019. Ln(Density) is computed by adding up all pupils in primary schools within 2km of the pupil's primary school. Local Wealth index is geospatial data provided openly by Meta corporation.

Table 2 Estimates on density and learning

Panel A: Reading					
N = 646914	[1]	[2]	[3]	[4]	[5]
Ln(Density)	0.062**	0.017**	0.018**	-0.025**	-0.037**
	[20.2]	[6.7]	[6.0]	[-3.6]	[-5.3]
Grade 2 score	Yes	Yes	Yes	Yes	Yes
Gender, Language, Home SES Index		Yes	Yes	Yes	Yes
School inputs			Yes	Yes	Yes
Agglomeration proxies:					
School size and local school density				Yes	Yes
Local wealth					Yes
R-squared	0.381	0.401	0.411	0.413	0.414
Panel B: Mathematics					
N = 646914	[1]	[2]	[3]	[4]	[5]
Ln(Density)	0.055**	0.012**	0.013**	-0.026*	-0.042**
	[14.8]	[3.2]	[2.9]	[-2.5]	[-3.8]
Grade 2 score	Yes	Yes	Yes	Yes	Yes
Gender, Language, Home SES Index		Yes	Yes	Yes	Yes
School inputs			Yes	Yes	Yes
Agglomeration proxies:					
School size and local school density				Yes	Yes
Local wealth					Yes
R-squared	0.328	0.347	0.361	0.364	0.365

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. SEs are corrected for spatial clustering at 2km distance. All models control for cohort effects. Cohorts refer to pupils who were in second grade in 2009, 2010, 2012 and 2013, and in 8th grade in 2015, 2016, 2018 and 2019. All columns in both panels are estimated with the same sample. Local school density refers to the logarithm of the number of primary school within two km of the pupils' current

Table 3 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 4 Summary statistics for 16-18 year old rural-urban migrants (census data)

	Treatment n = 50890		**Control** n = 13785		Difference	**Rural stayers** n = 451909
	Mean	S.E.	Mean	S.E.	z/t-stat	Mean
Child outcomes:						
Studied beyond primary	0.923	0.267	0.898	0.302	9.35	0.795
Correct age for grade	0.614	0.487	0.583	0.493	6.58	0.416
Secondary graduate	0.497	0.500	0.473	0.499	5.01	0.305
Studying beyond secondary	0.144	0.351	0.13	0.336	4.16	0.046
Age	17.00	0.82	16.94	0.81	0.72	16.89
Female	0.487	0.500	0.493	0.500	-1.2	0.447
Mother:						
Mother's age	42.54	6.7	42.36	6.77	0.75	44.26
Mother beyond primary	0.361	0.48	0.346	0.476	3.13	0.152
Mother beyond secondary	0.104	0.305	0.100	0.300	1.23	0.025
Mother speaks Spanish	0.647	0.478	0.657	0.475	-2.07	0.533

Notes: Treatment/Control 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 moved at most 5 years ago. 'Rural stayers' are the equivalent age group in rural districts whose parents have not moved.

Table 5 Time spent in urban environment and school attainment (census data)

	[1] Beyond Primary	[2] Correct grade for 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. Linear probability models. Robust standard errors in brackets. 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.

Appendix

Table A1 Full results for Reading

	[1]	[2]	[3]	[4]	[5]
			Grade 8 Reading		
Ln(Density)	0.059**	0.017**	0.018**	-0.025**	-0.037**
	[20.2]	[6.7]	[6.0]	[-3.6]	[-5.3]
Grade 2 reading score (z)	0.566**	0.524**	0.511**	0.507**	0.506**
	[106.0]	[134.3]	[149.1]	[148.7]	[149.4]
Female		0.042**	0.041**	0.041**	0.041**
		[12.2]	[12.1]	[11.8]	[11.6]
Spanish speaker		0.064**	0.080**	0.084**	0.077**
		[5.8]	[7.7]	[8.2]	[7.6]
Home SES Index		0.194**	0.149**	0.139**	0.133**
		[26.0]	[34.7]	[33.6]	[32.7]
PTR primary			0.000	-0.001	0.000
			[0.6]	[-1.4]	[-0.8]
PTR secondary			0.000	-0.002**	-0.002**
			[0.8]	[-3.7]	[-3.4]
% tenured primary			0.073**	0.053**	0.051*
			[3.3]	[2.6]	[2.5]
% tenured secondary			-0.191**	-0.241**	-0.254**
			[-10.4]	[-12.8]	[-13.5]
Private primary			0.307**	0.293**	0.286**
			[17.6]	[17.0]	[16.3]
Private secondary			-0.149**	-0.108**	-0.115**
			[-9.4]	[-6.3]	[-6.8]

Internet primary			0.020**	0.012+	0.01
			[2.9]	[1.7]	[1.4]
Internet secondary			0.067**	0.052**	0.045**
			[5.6]	[4.5]	[3.9]
Computers primary			0.080**	0.071**	0.068**
			[10.6]	[9.9]	[9.4]
Computers secondary			0.023+	0.003	-0.001
			[1.8]	[0.3]	[-0.0]
Agglomeration proxies:					
Ln(Size) primary				0.014**	0.005
				[3.0]	[1.1]
Ln(Size) secondary				0.064**	0.063**
				[10.9]	[10.9]
Ln(# Primary Sch. < 2km)				0.046**	0.044**
				[4.6]	[4.5]
Local wealth (z)					0.056**
					[7.1]
Observations	646914	646914	646914	646914	646914
R-squared	0.381	0.40	0.411	0.413	0.414

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. SEs are corrected for spatial clustering at 2km distance. All models control for cohort effects. Cohorts refer to pupils who were in second grade in 2009, 2010, 2012 and 2013, and in 8th grade in 2015, 2016, 2018 and 2019. All columns in each panel are estimated with the same sample.

Table A2 Full results for Mathematics

	[1]	[2]	[3]	[4]	[5]
			Grade 8 Mathematics		
Ln(Density)	0.053**	0.012**	0.013**	-0.026*	-0.042**
	[14.8]	[3.2]	[2.9]	[-2.5]	[-3.8]
Grade 2 mathematics score (z)	0.526**	0.494**	0.488**	0.485**	0.483**
	[84.2]	[93.1]	[95.2]	[95.2]	[97.1]
Female		-0.083**	-0.086**	-0.087**	-0.088**
		[-19.5]	[-20.4]	[-20.2]	[-20.3]
Spanish speaker		0.057**	0.073**	0.079**	0.070**
		[4.5]	[6.0]	[6.5]	[5.8]
Home SES Index		0.179**	0.126**	0.114**	0.107**
		[23.2]	[26.0]	[23.3]	[22.8]
PTR primary			0.001**	0.001	0.001*
			[3.3]	[1.4]	[2.0]
PTR secondary			0.001	-0.003**	-0.003**
			[0.7]	[-3.2]	[-3.0]
% tenured primary			0.008	-0.015	-0.018
			[0.4]	[-0.7]	[-0.8]
% tenured secondary			-0.234**	-0.291**	-0.307**
			[-9.2]	[-11.2]	[-11.9]
Private primary			0.315**	0.296**	0.287**
			[16.3]	[15.5]	[14.8]
Private secondary			-0.160**	-0.108**	-0.116**
			[-7.4]	[-4.5]	[-4.9]
Internet primary			0.021*	0.014	0.011

			[2.1]	[1.3]	[1.0]
Internet secondary			0.068**	0.049**	0.041*
			[4.1]	[2.9]	[2.5]
Computers primary			0.050**	0.040**	0.035**
			[5.6]	[4.6]	[4.1]
Computers secondary			0.028+	0.004	0
			[1.7]	[0.3]	[-0.0]
Agglomeration proxies:					
Ln(Size) primary				0.015*	0.004
				[2.4]	[0.7]
Ln(Size) secondary				0.076**	0.076**
				[9.3]	[9.3]
Ln(# Primary Sch. < 2km)				0.039*	0.037*
				[2.6]	[2.5]
Local wealth (z)					0.068**
					[6.4]
Observations	646,914	646,914	646,914	646,914	646,914
R-squared	0.328	0.347	0.361	0.364	0.365

Notes: **: p <.01, *: p <.05, +: p <.10. T-statistics in brackets. SEs are corrected for spatial clustering at 2km distance. All models control for cohort effects. Cohorts refer to pupils who were in second grade in 2009, 2010, 2012 and 2013, and in 8th grade in 2015, 2016, 2018 and 2019. All columns in each panel are estimated with the same sample.

Appendix: Within-family estimation of the effects of urban exposure

In this 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:

(A1) $Education_{ic} = \alpha + \beta AgeDif_{ic} * OrigRurality_d + \delta AgeDif_{ic} + \theta Fem_{ic} + \sigma_a + \gamma_{family} + \varepsilon_{ic}$.

The 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. The 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 for age. In the full sample, only 75% of pupils are in the correct grade for 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 A3 shows the summary statistics for the relevant sample, compared to all urban residents between the age of 7-18.

Table A4 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 A4, the family fixed effects are removed. Interestingly the effect of urban exposure on timely school progression is about a 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 A5 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.

Table A3 7-18 year old urban residents who lived in a different location 5 years ago, censuses 2007 and 2017 (census data)

	Estimation sample: Recent migrants 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 for 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 A4 Effect of exposure to urban area (census data)

Dependent: Correct grade for 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 A5 Effect of exposure to urban area - alternative estimations (census data)

	[1]		[2]	
	Correct grade for 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.