Human Capital and Aggregate Income Differences in Mexico

Jorge N. Valero-Gil

Facultad de Economía, Universidad Autónoma de Nuevo León

Magali Valero[1]

University of Michigan at Dearborn

This Version: December 2020

This article has been accepted for publication in *Applied Economics*, published by Taylor and Francis.

Abstract

We study the relationship between differences in human capital and differences in product per worker of the federal entities of Mexico. We consider both quantity and quality of education in human capital formation, and use two methods for aggregating these two dimensions of education: a multiplicative and an additive model. Our measures of quality of education are constructed using the OECD's Program for International Student Assessment (PISA) math achievement test scores in the additive model, and using the differences in the returns to education of the states in the multiplicative model. Our results are consistent to different methodologies and data sources. We find that variations in human capital explain upwards of 40% of the variations in state GDP per worker. Our results indicate that Mexican states should place more emphasis both in the quantity as well as quality of schooling, in order to improve the living standards of their population.

JEL Codes: I25, I26, J24, R11, O54, O15

Keywords: Education; Human Capital; Mexico; Development Accounting

[1] Corresponding Author. 19000 Hubbard Drive, Dearborn, MI, 48126. Phone: (313) 583-6532. Email: mvalero@umich.edu.

Human Capital and Aggregate Income Differences in Mexico

1. Introduction

This paper studies differences in human capital, and how these impact economic development of the federal entities in Mexico[2]. Significant differences exist between the Mexican states, in terms of GDP per worker and GDP per hour worked, as well as in terms of human capital measured through the average years of schooling. In this paper we study the role of human capital as a possible source of income variations across the Mexican states. We refer to human capital as the one derived from schooling, and not as the one derived from work experience as in Bils and Klenow (2000). We first study the differences in human capital across states and then study whether differences in human capital across states can explain differences in GDP per worker and per hour worked.

International empirical studies suggest that differences in human capital (in terms of both quantity and quality of education) explain between 20 and 40 percent of the income differences of countries (Schoellman (2012), Hanushek and Woessmann (2012b)). Within a country, Hanushek, Ruhose and Woessmann (2017) find that between 20-30 percent of the variation in a state's GDP per capita can be explained by human capital differences in the case of the United States.

We consider both the quantity and the quality of human capital. We estimate human capital using two models. In the additive human capital formation model (Hanushek et al. 2017) the effects of years of schooling are added to its quality. We follow Hanushek et al. (2017) who utilize test scores to measure the quality of education, we use the PISA

[2]The federal entities include the 31 Mexican states in addition to Mexico City, we refer to these as the 32 Mexican states.

Mathematics test scores as a measure of quality of schooling. In the multiplicative human capital formation model (Schoellman, 2012) years of schooling and quality of schooling are multiplied in the human capital production function. Following Schoellman (2012) we use the returns to education to approximate the quality of education, and like Card and Krueger (1992), we attribute the differences in returns to education to differences in quality of education in each of the Mexican states.

Mexico offers a perfect framework for studying the effects of quantity and quality of education in the formation of human capital and the participation of such in GDP per worker. Starting in 2003 and every three years since then, the PISA test is representative for each of the 32 Mexican states. An additional advantage of studying the case of Mexico is the composition of its GDP as compared to that of the United States. Hendricks (2002) finds total factor productivity differences cannot explain income differences between the U.S. and Mexico; hence differences between the two countries can be attributed to differences in either physical or human capital. Moreover, Hall and Jones (1999) suggest differences in human capital between Mexico and the U.S. are more important than differences in total factor productivity or in physical capital in terms of explaining the differences in output per worker of the two countries. This suggests that human capital is an important component in explaining income differences between Mexico and the United States.

The study of a single country also allows us to control for other factors that might affect income, hence we can estimate the effects of quality of schooling to income with greater certainty. Prior literature focuses on explaining the portion of income differences between countries that can be attributed to human capital. In these studies, the country where the individual is working might be different to the country where the education was received. By focusing on a single country, although the state where the education was received and the state where the individual works might be different, we are able to control for language and the culture of work, which one expects vary less within a country than internationally.

It is possible that in the case of Mexico human capital differences among the states are important in explaining income differences. Bils and Klenow (2000) model the effects of education on growth of GDP per capita, and find that schooling can explain less than one third of the schooling/growth cross-country relationship. They use parameters that assume decreasing returns to education, based on estimates by Psacharopoulos (1994) for several countries. Even though decreasing returns to education arises from the comparison between countries, this does not imply that each country has decreasing returns. By assuming diminishing returns, the importance of human capital in their model falls as more human capital is acquired, but growth in human capital can be important in explaining the schooling/growth relationship if there are no diminishing returns to education. Harberger and Guillermo-Peón (2012) find that in the case of Mexico the returns to education are not decreasing[3], which would imply a stronger importance of schooling in explaining income differences in such case.

The study is timely and important in the case of Mexico. The current public education system was implemented in 1959 and had not had any significant changes since then. A proposal of an educational reform which looked to improve the quantity and quality of education in Mexico was presented in late 2012 by then president Enrique Peña Nieto, and was subsequently signed into law. However, in 2019 the educational reform was repealed. The quality of education has restrictions and differences at the state level in part due to the presence of large unions in the Mexican education sector, and to its management by the Mexican state.

[3] Patrinos, Ridao-Cano and Sakellarious (2006) find increasing returns to education for other countries also used in Psacharopulos's (1994) sample.

Implicit in the educational reform proposal is the key assumption that such improvements in education will lead to a reduction in inequality and to greater economic development of the country, which will lead to an improvement in the lives of the country's constituents. There is some empirical evidence that a higher quality of education is related to higher wages in the case of Mexico. For instance, De Hoyos, Estrada and Vargas (2018) find a positive relationship between individual test scores and individual wages. Further, their findings indicate that higher test scores are associated with a higher probability of a student going to college. The study of human capital as a possible source of income differences among the Mexican states is thus significant for the country.

Our findings indicate that quality adjusted human capital explains upwards of 40% of the variations in income per worker of the states. This suggests that human capital is a significant component of income differences between the states in Mexico. As a comparison, Hanushek et al. (2017) find that between 20-30% of the variation in GDP per capita can be explained by human capital differences in the case of the United States. This study is important as there are no attempts in the literature to study the role of human capital as a source of income differences among the Mexican states. Further, only one other work studies the role of human capital as a source of income differences among states in a country (Hanushek et al., 2017 for the United States) where, as mentioned above, human capital seems to play a more important role in the case of Mexico. This study allows us to understand the role of human capital on the income of the people of the Mexican states, so that states can approve public policies effective in improving the income of their constituents.

The paper is structured as follows. In Section 2 we discuss the sample selection and data. Section 3 presents the analytical framework. Section 4 describes our measures of quality of education and in Section 5 we use those measures in addition to quantity of schooling in forming human capital. Section 6 presents the decomposition of variations in GDP that are accounted by differences in human capital. In section 7 we evaluate the robustness of the results. Section 8 discusses the results and Section 9 concludes.

2. Sample selection and Data

To estimate the working population and the hours worked in the labor market we use the 2010 Census (Censo de Poblacion y Vivienda, INEGI (2012)), the available data includes more than 11.9 million observations. Following Hanushek, Ruhose and Woessmann (2017) we select the working population between the ages of 20 and 65 who are not currently in school, leaving 3,304,715 observations, which, using expansion factors, represent 36.3 million workers. An alternative source of data is the National Employment Survey, ENOE (Encuesta Nacional de Ocupacion y Empleo). We corroborate the robustness of our results by using data from the 2016 ENOE in section 7 of this study. Taking the population between 20 and 65 years of age who are declared working and not currently in school leaves 136,197 ENOE observations representing 42.5 million workers.

To estimate state GDP per capita we use INEGI (2018). GDP per capita and GDP per hour worked are found by taking the state GDP and dividing by the selected working population and by the annual hours worked of the selected population, respectively, using either the 2010 Census or data from the third trimester of the 2016 ENOE. Table 1 shows GDP and years of schooling by state. The first two columns show GDP for 2010 and 2016, measured in millions of 2013 Mexican Pesos, the difference between them implies an annual growth rate of 2.89% on average. There are significant income differences among the 32 Mexican states. For instance, in 2016 the GDP per worker of Coahuila was 546,131 Mexican Pesos (MP), more than twice that of Michoacán, which was MP 261,831. Excluding the two states where the production of oil occurs in Mexico, Campeche and Tabasco, the second largest 2016 GDP per worker is Nuevo Leon at MP 642,637; which is more than three times higher than that of the second lowest which is Oaxaca at MP 195,321. The standard deviation in state incomes (excluding Campeche and Tabasco) is MP 146,715, which is higher than 39 percent of the national average. As a comparison, Hanushek et al. (2017) report the 2007 standard deviation in state incomes being around 15% of the national average in the case of the United States. The last two columns of Table 1 show the average years of schooling for the population in the work force in years 2010 and 2016. The difference in years of schooling between the top and bottom states is four years or five standard deviations. In comparison, the difference in years of schooling between the U.S. states with the maximum and minimum values is 3.6 standard deviations. Hence, in addition to significant income differences, we also observe significant differences in terms of years of schooling among the Mexican states.

3. **Analytical Framework**

To measure the contribution of human capital differences on income, we start with the production function $Y_i = K_i^{\alpha} \left(A_i H_i \right)^{1-\alpha}$, where Y is income, K is capital, A is total factor productivity, H is the amount of human capital-augmented labor used in production, and α refers to the proportion of income allocated to capital. This function allows the decomposition of the variations in product per worker into the different factors including human capital per worker *h*, as in Hall and Jones (1999) and Hanushek et al. (2017) among others. The production function on a per worker basis can be written as:

$$\frac{Y}{L} \equiv y = h * A * \left(\frac{k}{y} \right)^{\alpha/(1-\alpha)} \qquad (1)$$

where $k \equiv \frac{K}{L}$ is the relation of capital to labor, and y is income per worker, We can decompose the variance from log GDP per worker as in Klenow and Rodríguez-Clare (1997). This decomposition is presented by Hanushek, Schwerdt, Wiederhold and Woessmann (2015), and by Hanushek et al. (2017) as:

$$\frac{\operatorname{cov}\left(\ln(y),\ln(h)\right)}{\operatorname{var}\left(\ln(y)\right)}+\frac{\operatorname{cov}\left(\ln(y),\ln\left(\left(\frac{k}{y}\right)^{\alpha/(1-\alpha)}\right)\right)}{\operatorname{var}\left(\ln(y)\right)}+\frac{\operatorname{cov}\left(\ln(y),\ln(A)\right)}{\operatorname{var}\left(\ln(y)\right)}=1 \tag{2}$$

The first term refers to the variation in GDP per capita that can be attributed to human capital differences, which is the focus of this study. As a robustness check, following Hall and Jones (1999) we calculate the 5-state measure to account for the contribution of the difference in human capital to the difference in GDP per capita between the 5 richest and 5 poorest states. Additionally, we redo the measurement using the 3 richest and the 3 poorest states (3-state measure).

3.1 Quality-adjusted Human Capital Measures

In addition to using years of schooling to measure human capital, we also account for the quality of schooling across states. We consider two models that explain human capital formation as a function of the years of schooling, s, and the quality of schooling, Q. The first is an additive model, used by Bils and Klenow (2000) and Hanushek et al. (2017) among others; in the additive model the years of schooling measure is added to the quality of schooling measure in the human capital formation function. The second model is a multiplicative model, designed by Schoellman (2012), where years of schooling and quality of schooling are multiplied in the human capital production function.

3.1.1 Additive human capital formation

In the additive model human capital formation is formulated as:

$$h = e^{\theta s + \pi Q} \tag{3}$$

The earnings gradients to years of schooling θ, and quality of schooling π, establish the relationship of quantity and quality of schooling with human capital, *h*. There are two issues when determining the earnings gradients for years of schooling (θ) and quality of schooling (π). The first is that the obtained values are useful in determining human capital and wages through the working life of individuals, and that they include individuals already in the workforce. The second is that the values of θ and π have to be estimated simultaneously, otherwise θ could contain information about the quality of schooling and the cognitive abilities of the individual.

Our approach is to take the parameters of the earnings gradients from the current literature. The most common way of measuring the schooling gradient, θ, is using Mincer regressions as in Card and Krueger (1992) and Shoellman (2012). However, the exclusion of cognitive skills measures confounds the estimation and hence the estimation is not appropriate in our context. We look for joint estimates of the earnings gradients for years of schooling and quality of schooling.

The quality factor π is hard to estimate, as we look to estimate the effect of quality of schooling once the individuals are working. One possible way is to estimate returns early in an individual's career, possibly omitting differences across lifetime earnings. Lazear (2003) and Murnane, Willet, Duhaldeborde and Tyler (2000) observe individuals when they are in school and then again years later once they are in the labor force. They then relate differences in earnings to differences in education and abilities. They find that a one standard deviation in mathematics performance when in school leads to a 9-15% increase in earnings. An alternative is to use estimates of lifetime earnings based on skills that are measured during the workers career. For instance, Hanushek et al. (2015) use data from the “Programme for the International Assessment of Adult Competencies” (PIAAC) find ranges of the earning returns to cognitive skills of 0.14 to 0.28 for the 23 countries in their sample. Their estimates for the United States are 8.1% return to school attainment and 13.8% return to cognitive skills. Hanushek and Woessmann’s (2012a) estimate of π is 14%.

Facing the problem of no available joint estimates for Mexico we select the parameters that best fit our purpose given the limitation faced. Hanushek and Zhang (2009) estimate the value for individual literacy scores to school attainment and provide joint estimates for the parameters for 13 countries. In their baseline model, a classical Mincer equation with years of schooling, the estimate for returns to education is 11% for the United States and 11.3% for Chile (the only country in the sample from Latin America), with the returns in other countries being much lower (below 8.4%). Given that our estimate of returns to education for Mexico using Mincer equations is 10.4%[4], our approach is to use the values in the literature for the United States and use the parameters estimated for Chile as a robustness check. Once Hanushek and Zhang (2009) make adjustments cognitive skills, their estimation of θ is 8.0% for the United States and 8.9% for Chile, while their estimations of π are 19.3% and 13.1% respectively. To make our study comparable to that of Hanushek et al. (2017) who study how differences in human capital can explain differences in income in the United States, we follow by using values of 8.1% for θ and 17% for π. As a robustness check, in Section 6 we vary the

[4] In unreported results. The regression specification is similar to that in equation (5), without the dummies for states j.

value of π to 13.1%, which is the value estimated for Chile by Hanushek and Zhang (2009) and find similar results.

3.1.2 Multiplicative human capital formation

Schoellman (2012) defines human capital as:

$$h\left(s_j, Q_j\right) = \exp\left[\left(s_j Q_j\right)^{\eta} \Big/ \eta\right] \tag{4}$$

This equation interacts, for a migrant in the United States from country j, the number of years of schooling s_j with the quality of schooling Q_j, where the exponent η moderates the interaction between both variables. Schooling quality and years of schooling are positively correlated as long as $0 < \eta < 1$. Because Schoellman (2012) associates the quality of schooling Q to the rate of return to education of migrants of country j in the United States, $s_j Q_j$ would be the percent increase in earnings from education for a person with years of schooling s and quality Q_j, like in the Mincer equation. We follow Schoellman's methodology for estimating quality adjusted schooling, applying it to each of the Mexican states in Section 5.

4. Measures of Education Quality

4.1 Rates of Return to Education

In the multiplicative human capital formation model, quality of schooling Q_j (in equation 4) is estimated using returns to schooling by state, which we estimate as θ_j through the following:

$$\ln y_{ijk} = \alpha + \delta_j + \mu_k + \theta_j s_{ijk} \lambda_j + \gamma X_{ijk} + \varepsilon_{ijk} \tag{5}$$

Where *i* denotes individual, *j* denotes the state of birth of the individual, and *k* the state in which the individual was residing and working at the time of the 2010 census. The census does not provide data on where the individual received their education, therefore, the implicit assumption we make is that the individual received its education in their state of birth. For the regression, we select from our sample individuals who work 30 or more hours per week and earn at least $500 pesos per month, who were born in a Mexican state and for which we have data on their years of schooling, leaving 2,244,341 observations. We regress, for 2010, the log of hourly income against years of schooling (s_{ijk}) and a vector of controls X_{ijk} that includes experience (experience = age – education – 6) to the 1-4th powers[5], sex, and size of the city (under 15,000 habitants, 15,000 – 100,000 habitants, and more than 100,000 habitants), and if the individual is a migrant into the state or not. The vector δ_j refers to the parameters of the state of birth fixed-effect and μ_k to the vector of parameters representing the state of residence. Additionally, λ_j is a dummy variable for the population born in state j, while θ_j is the vector of parameters that represent the state-specific return to schooling. By controlling for the state of birth, we control for differences in education and in the learning habits of the state of birth, and by controlling for the state of residence we control for differences in technology that the states could have. For instance, Atkin (2016) finds that the manufacturer export sector in Mexico generated an abundance of low-ability work opportunities in exporting zones; the dummy variable would then control for these effects which occurred mainly in the north of Mexico. Under this specification, we allow the return to schooling to vary by state by interacting years of schooling s_{ijk} with an indicator variable for the state of birth λ_j which is presumably where the individual received his education.

---

[5] We include experience to the 1-4th powers following Murphy and Welch (1990), as justified in Hamlen and Hamlen (2012).

4.2 PISA Mathematics test achievement scores

We use the OECD's Program for International Student Assessment (PISA) math achievement test scores as an additional measure of cognitive skills. Starting in 2003 the PISA test is representative for each of the Mexican states. Mathematics was the major subject of PISA in 2003 and 2012. Mexico scored well below the OECD average of 494 in the Mathematics portion of the 2012 PISA test (OECD 2014), with a score of 413, close to that of other Latin American countries and well below the score for the United States (481). According to the OECD, Mexico placed the equivalent of 2 years of schooling below the average OECD countries for same-grade students, and about 1.6 years of schooling lower than the United States.

We take the mathematics score on the PISA test for each state for the years 2003, 2006, 2009 and 2012[6]. We then calculate the average score across states by year, and normalize it with mean 500 and standard deviation of 100, to make it comparable to the data in Hanushek, Ruhose, and Woessmann (2017). We use the 2003-2012 state average as a measure of cognitive abilities of the working population[7]. We assume that test scores (and therefore quality) are stable over time, even though test scores can vary across successive tests[8]. We also assume that the average PISA test scores apply to the working population in 2010 and 2016.

---

[6] The exam was not administered in the state of Michoacán in 2003 and 2012, and it was not administered in Oaxaca or Sonora in 2012. We replace the 2006 value for the 2003 missing value and use the 2009 value for the 2012 missing values.

[7] We use the average score across time to avoid sample differences which occur each year the exam is administered, as these differences can be significant. For instance, in the state of Mexico the score was 399 in 2003, 433 in 2006, 415 in 2009 and 413 in 2012.

[8] We do not have test results at the state level prior to 2003, and it is difficult to infer any trends from the PISA test results we do have. As an example, the overall PISA mathematics score for Mexico was 387 in 2000, 385 in 2003, 405 in 2006, 418 in 2009, 413 in 2012 and 408 in 2015. One could erroneously infer an improving trend by looking at the 2000-2009 scores, which clearly is not the case once we see 2012-2015 scores.

Table 2 shows the standardized PISA mathematics test scores for 2003 and for the average 2003, 2006, 2009 and 2012. The state with the highest average standardized score is Nuevo Leon, with 704, the lowest scoring state is Chiapas with 279. These two test scores are four standard deviations away from each other. As a comparison, Hanushek et al. (2017, online appendix Table 2) report the biggest difference in average standardized NAEP scores is between Minnesota and Missouri, 534.8 and 450.8 respectively, a difference of less than one standard deviation. These significant differences in quality of schooling across Mexico mirror the significant variations in GDP and in years of schooling discussed in Section 2.

We next adjust mean test scores for each state for interstate migration, the autoselection of interstate migration, and for international migration also considering its possible autoselection. We follow the methodology in Hanushek et al. (2017). The adjustments are summarized in Table 3. Adjusting for interstate migration, the mean PISA test score is unchanged while the standard deviation falls to 88. After adjusting the tests scores of state residents for their educational background, the mean falls indicating negative autoselection (individuals with less schooling). Finally, when we adjust for international migration the difference between the maximum and minimum state scores is at its lowest, with a difference in scores of 304 points, which represents more than 3 standard deviations[9].

5. Incorporating Schooling Quality into Human Capital

We use two approaches to incorporate the measures of schooling quality into a human capital production function: the multiplicative and the additive model of human capital.

[9] The appendix (available from the authors) details the methodology used, and shows the adjustments of test scores, the returns to schooling and the human capital estimates using the additive and multiplicative models by state.

5.1 Estimating human capital using the multiplicative model.

Once we have the estimates of quality of education for each state, we still need the parameter η in order to account for quality in human capital formation through equation (4). We follow Schoellman's (2012) model and procedure for estimating η[10]. The value for η is derived from the estimation of the elasticity of years of schooling with respect to education quality, η/(1 – η), which we get through the following regression:

$$\ln s_j = c + \frac{\eta}{1-\eta}\ln\left(Q_j\right) \tag{7}$$

Where s_j is years of schooling in state j; and quality of schooling Q_j is estimated with returns to schooling θ_j in this equation and in equation (4). From this estimation of elasticity, one can infer the value of η to be used in equation (4). As part of this estimation, Schoellman (2012) recommends using standardized achievement test scores as instruments for the returns to schooling, as it is an additional measure of the quality of schooling which is unlikely to affect the years of schooling. Besides, one would not expect reverse causality of the test scores which are administered to 15-year-old students towards the years of schooling of the state. The instruments we use are the average adjusted PISA test scores and the distance from the state capital to the closest U.S. border city. We include distance to control for changing returns to education once trade was opened in Mexico. Open trade and hence the development of the export and maquiladora industry starting in 1986 could have had an impact in the returns to education for individuals in those industries, which are mainly located near the Mexican border with the U.S. Although Schoellman (2012) shows that a higher rate of return can affect years

[10] We follow equation (9) of Schoellman (2012).

of schooling, Atkin (2016) finds that for those closer to the U.S. border years of educational attainment fell with the arrival of new export-manufacturing jobs.

Table 4 shows the estimates of the elasticity of school attainment and of η. The values of the elasticities are 0.24 in the OLS model and 0.25 in the model that uses instruments. Our elasticities are lower than those estimated by Schoellman (2012), which are as high as 1.36. To prove if the instruments are weak, Bound, Jaeger and Baker (1995) suggest the report of F and the partial R squared of the first stage estimates. We find a partial R squared of 0.62 and F higher than 22, which exceeds the value of 11.59 suggested by Stock, Wright, and Yogo (2002) to make the instruments reliable. Further, the Wald test of Stock and Yogo (2005) shows that at the 5% level we can have a rejection rate of no more than 10, therefore the instruments are not weak. We can reject the null hypothesis that the independent variable is endogenous.

We use the estimate for η and equation (4) to estimate human capital, the estimates show the top and bottom states are about 4 standard deviations away, indicating large dispersion in our human capital estimates among the Mexican states.

5.2 Estimating human capital using the additive model.

Hanushek et al. (2017), model human capital h(s, Q), as a function of years of schooling and of cognitive abilities of the working population: $h = e^{\theta s + \pi Q}$. The value of s for each state is determined by the average years of schooling for the working population. Setting the value of π at 17%, we estimate human capital according to equation (3). We use the average adjusted PISA Math test scores as a measure of quality of schooling. We standardize the values with mean zero and variance of one, then further adjust so that our relative measure of quality (standardized difference in the PISA test score) has the minimum value of zero.

We estimate the Kendall Rank correlation for the estimates of human capital. We find a high correlation between human capital estimates in the case of the multiplicative model when using OLS or instruments (both generate the same order of the states for human capital). Considering the estimates with instruments, the additive and the multiplicative model order human capital in the states in the same way in more than 52% of the cases. Using Pearson correlation coefficients, the correlation between human capital in the additive and multiplicative model (with instruments) is greater than 70%.

6. Decomposing state variations in GDP into contributions accounted by differences in quality and quantity of human capital

We next decompose the variation in product that can be attributed to differences in human capital. We exclude states whose industry structure makes GDP unlikely to be described well by a capital and labor production function, hence, we exclude those with abundant natural resources following Hall and Jones (1993) and Hanushek et al. (2017). There are two states where we cannot expect a direct relationship between human capital and product per worker, these are the states were the production of oil occurs in Mexico: Campeche and Tabasco. According to the Economic Census of 2008 (INEGI, 2015), 96% of the value added in Campeche and 82% in Tabasco corresponds to the extraction of oil, thus one would expect a very weak relationship between GDP per worker and human capital in these two states, they are excluded from our sample at this point.

Table 5 shows the percentage of the variability in income attributed to human capital, estimated as in equation (2). In the additive model years of schooling explain 18% of the variation in GDP per worker, while cognitive abilities explain 31% using the adjusted PISA test scores. Therefore, human capital differences explain 49% of the differences in GDP. These estimates use values of 17% for π and 8.1% for θ. Alternatively, using π=13.1%, cognitive abilities explain 24% of the variation in income; and human capital explains 42% of the variations in GDP.

The bottom portion of the table refers to the multiplicative model. In this case human capital explains 70% of the variations in GDP per worker. This estimate is robust to using only adjusted test scores as an instrument.

Following Hall and Jones (1999) we also compare the relation between human capital in states with the highest and lowest GDP per hour worked in the last four columns of Table 5. The five-state and three-state measures provide similar results. Human capital differences explain between 54% and 77% of the differences between GDP per capita of the richest and poorest states when using the additive model, and between 65% and 86% when using the multiplicative model. Both models maintain their strong predictive power even for the case of largest and smallest states in terms of GDP per hour worked. Therefore, these estimates of the variation in income attributed to human capital do not fall once we use the five-state or three-state measure. A potential explanation is that we excluded the two outliers: Campeche and Tabasco. The results when we include Campeche and Tabasco will be discussed in Section 8.

7. Robustness Checks

We now study if our estimates are robust to changes in a) the test period and data source; 2) using migrant workers only; and 3) using PISA 2003 test scores instead of the 2003-2012 averages.

7.1 Results for 2016

Instead of using the 2010 Census, we could use the 2016 ENOE (Encuesta Nacional de Ocupacion y Empleo- National Survey of Occupation and Employment), which provides quarterly data on the working characteristics of the population. The more recent data comes at the cost of a much smaller sample. We use the third quarter from ENOE for 2016 (2016 is the latest year with GDP data by state). Imposing the same data restrictions as before our sample is 82,845 observations (compared to 2,244,341 with the Census). The Census indicates place of birth, and in the case of international migration country of birth, allowing the adjustment of PISA test scores for international migration. This information is not available when using the ENOE.

The results using GDP per hour worked[11] are shown in Table 6, Panel A shows the additive model and Panel B the multiplicative model. The percentage explained by human capital, which was 49% or more using the 2010 Census (Table 5), falls to 46%. The proportion of the variation in the GDP per hour explained by human capital in the multiplicative model, which was 70% using the 2010 Census, is now reduced to 41% with the 2016 ENOE. A possible reason for the strong fall is that the national rates of return to education, which were estimated at 10.2% using the 2010 Census, are 7.2% using the 2016-III ENOE[12].

Overall, our findings indicate that using ENOE 2016 the variations in the log of human capital still explain between 40% (multiplicative model) and 50% of the variations in log GDP.

[11] The results using GDP per worker are qualitatively and quantitatively similar.
[12] The ENOE sample is much smaller and income measurement more limited compared to the Census. Using hourly wages from the National Survey of Income and Expenditures of Households (ENIGH –Encuesta Nacional de Ingresos y Gastos de los Hogares de Mexico) we estimate national rates of return to education of 10% for both 2010 and 2016. Unfortunately, this survey is not representative at the state level.

7.2 Using only migrants in the multiplicative model

A possible issue with the study of how human capital affects income is the problem of causality between income and education, that is, it is possible that the richest states will not only provide better education but also better returns to education. To control for this possibility, we separately study a subsample of individuals who are migrants into the state. To measure the returns to education of a particular state, we take all individuals born in that state (assuming they were educated there), but that later moved to a different state. Hence, we are taking individuals educated in one state and observing their incomes in other states. This way we break the problem of causality of income to education, as migrants are taking their education to a different state.

Other studies also use migrant earnings to study the importance of education (Hendricks, 2002) or its quality (Schoellman, 2012; Hanushek et al., 2017). They obtain data about migrants in their home country and observe the results on the productivity of the labor market in the United States. This is not without problems, as the U.S. market may not be the appropriate one. For instance, Schoellman (2012) estimates returns to education for migrants in the U.S. to use them in measurements of quality of education, but these returns can be very low, such as 1.8% for Mexico or 2.3% for Portugal, that they might not reflect the returns to education or quality of education of individuals in their home countries. Therefore, it is important to estimate returns to education in Mexico by using individuals educated and residing in Mexico, while breaking the causality from state income to education, even though biases in our estimations could arise given that the migrant population auto-selects (Borjas, 1987). When we consider the migrant population, the 2010 sample is 397,870 and the 2016 sample is 17,124 observations. Table 7 shows that variations in human capital explain 65% of the variations in GDP per hour worked in 2010 and around 55% in 2016, when using migrant workers.

7.3 Results using PISA 2003

We re-estimate the results using the 2003 PISA test scores in lieu of the 2003-2012 average in the additive model. A 15 year old student who took the test in 2003 would be 22 years old in 2010, and 28 in 2016. In this case, we assume that the 2003 score is the appropriate one that applies to the working population in 2010 and 2016. Our results are robust to this correction. Table 8 shows that under such scenario differences in the quality and quantity of education explain between 42% (using the 2016 ENOE) and 46% (using the 2010 Census) of the differences in the incomes (GDP per worker) of the states.

8. Discussion of Results

Hanushek et al. (2017, Table 2) find human capital explains 15% to 22.8% of the variations in GDP between the U.S states, after the sensitivity analysis these estimates are 18.1% to 31.5%. In Schoellman (2012, Table 2), human capital explains between 19% and 36% of the variations in income in a cross-country study. Our results, however, show that in the case of the Mexican states the variability in GDP that can be explained by human capital is much larger, upwards of 40%.

One possible explanation for the larger estimates for Mexico is that we excluded the two Mexican states characterized by their oil extraction activity: Campeche and Tabasco. The value of production in these states depends in large part on the price of oil and on the existence of oil reserves, and not on the amount of labor and capital, so empirically this would affect the value of total factor productivity. Table 9 shows results of our analysis with and without the inclusion of these two states. The variance in ln GDP per hour in 2010 is 0.9 when we exclude Campeche and Tabasco, 0.10 including Tabasco and 0.23 including both states. The variance in GDP increases more than 10-fold by including the two states, so variations in human capital are less able to explain this variance in GDP. By excluding them, we exclude a source of the variations in total factor productivity and the variance is better explained directly by the inputs. In particular, following the additive model of Hanushek et al. (2017) human capital explains 49% of the variations in income in 2010 excluding the two states but only 7% if they are included. Similarly, following Shoellman's (2012) multiplicative model human capital explains 70% of the variations in income in 2010 when the two states are excluded but only 28% when they are included.

Another possible explanation for why the variations in human capital explain such a large fraction of the variations in GDP among the Mexican states is that capital (K) is not good at explaining such variations in income. In particular, the variance in GDP per worker can be attributed to three components as indicated in equation (2), the portions due to variations in human capital, to variations in capital, and to variations in total factor productivity. There is no data available for capital by state, but there is data on the National Accounts on the gross fixed capital formation by state between 2003 and 2017 (INEGI, 2019). Using this data and the method of perpetual inventories, we find a negative covariance between income and capital (the second term in equation 2) for both 2010 and 2016. Therefore, there is room for human capital to explain the variations in income of the Mexican States[13].

[13] We follow the program in Amadou (2011) and use a depreciation rate of 5% and values for α of 1/3, ½ and 2/3 for both 2010 and 2016, and all give estimates of a negative covariance term. Given the short time series available, these results should be interpreted with caution.

9. Conclusions

Our study shows that differences in schooling, both in terms of quantity and quality, explain upwards of 40% of the changes in GDP per worker. This result is robust to taking the five states with higher and lower GDP per hour worked, and also to taking the top and bottom three states. We used two methodologies to measure quality adjusted human capital, in the additive model quantity and quality of schooling are added in the human capital production function, and quality of schooling is measured using the achievement scores of the PISA mathematics test. The variations in human capital explain between to 40-50% of the variations in GDP per worker. The result is robust to using a different survey and year for data, the 2016 ENOE in lieu of the 2010 census, and to using the 2003 PISA test results instead of the 2003-2006-2009-2012 average.

In the multiplicative model human capital explains 70% of the variations in GDP per hour worked. The result is not driven by a problem of causality from GDP to education, as the results still hold when we take individuals who were born and probably educated in a state different from their current state of work. When we use the 2016 ENOE, the estimates fall to the same range as the additive model, 40-50%. The reason for the fall is that the rates of return to education estimated are much lower when using 2016 ENOE compared to when we use the 2010 Census.

Of particular importance is which states are included. We exclude Campeche and Tabasco as they are oil producing states where GDP is not reflective of the amount of labor and capital of the state. When including these states variations in human capital explain only

7% of the variations in GDP per hour in the case of the additive model and 28% in the case of the multiplicative model.

A variation of the model in Blis and Klenow (2000) would show that the effects of education on growth of GDP per capita could be important in cases where countries do not exhibit diminishing returns to education, such as the case of Mexico. Our results show that human capital differences among the states are important in explaining income differences in Mexico.

References

Amadou, D. I. (2011). *STOCKCAPIT: Stata module to calculate physical capital stock by the perpetual-inventory method.* Statistical Software Components S457270, Boston College Working Papers in Economics.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in Mexico. *American Economic Review, 106*, 2046-2085.

Bils, M., & Klenow, P. J. (2000). Does Schooling Cause Growth? *The American Economic Review, 90*, 1160-1183.

Borjas, G. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review*, 77, 531-553.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90*(430), 443-450.

Card, D., & Krueger, A. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *The Journal of Political Economy, 100*, 1-40.

De Hoyos, R.; Estrada, R.; and Vargas, M. (2018). *Predicting Individual Wellbeing Through Test Scores: Evidence from a National Assessment in Mexico.* Policy Research Working Paper;No. 8459. Washington, D.C.: World Bank Group.

Elizondo Mayer-Serra, C. E. (2009). Perverse equilibria: unsuitable but durable institutions. In S. Levy & M. Walton (Eds.), *No growth without equity? Inequality, interests, and competition in Mexico* (pp. 157-202). New York: Palgrave Macmillan and the World Bank.

Harberger, A. C., & Guillermo-Peón, S. (2012). Estimating private returns to education in Mexico. *Latin American Journal of Economics, 49*, 1-35.

Hall, R. E., & Jones, C. I. (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others? *The Quarterly Journal of Economics, 114*, 83-116.

Hamlen, S. S., & Hamlen, W. A. (2012). The inconsistency of the quadratic Mincer equation: a proof. *Theoretical Economics Letters*, *2*(02), 115.

Hanushek, E. A., Ruhose, J., & Woessmann, L. (2017). Knowledge Capital and Aggregate Income Differences: Development Accounting for US States. *American Economic Journal: Macroeconomics*, *9(4)*, 184-224.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to Skills around the World: Evidence from PIAAC. *European Economic Review, 73*, 103-130.

Hanushek, E. A., & Woessmann, L. (2012a). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth, 17*, 267-321

Hanushek, E. A., & Woessmann, L. (2012b). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics*, *99*, 497-512.

Hanushek, E. A., & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital*, *3*, 107-143.

Hendricks, Lutz. 2002. "How Important Is Human Capital for Development? Evidence from Immigrant Earnings." *The American Economic Review*, *92*, 198-219.

INEGI. (2012). Censo de Población y Vivienda 2010. Retrieved July 13, 2018, from Instituto Nacional de Estadística y Geografía http://www.beta.inegi.org.mx/proyectos/ccpv/2010/

INEGI.(2015). *"Minimonografías por entidad federativa, censos económicos 2009"*. .Retrieved from http://www.inegi.org.mx/est/contenidos/espanol/proyectos/censos/ce2009/minimonografias.asp

INEGI. (2018). Producto interno bruto por entidad federativa. Base 2013. From Instituto Nacional de Estadística y Geografía. https://www.inegi.org.mx/sistemas/bie/?idserpadre=10200070#D10200070

INEGI. (2019). Cuentas Nacionales - Gobiernos estatales, formación bruta de capital fijo, base 2013. Retrieved 05/31/2019, from Instituto Nacional de Estadística y Geografía https://www.inegi.org.mx/sistemas/bie/

Klenow, P., & Rodriguez-Clare, A. (1997). The neoclassical revival in growth economics: Has it gone too far? *NBER Macroeconomics Annual 1997, Volume 12* (pp. 73-114): MIT Press.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review, 10*, 179-214.

Murnane, R. J., Willet, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management, 19*, 547-568.

Murphy, K. M., & Welch, F. (1990). Empirical age-earnings profiles. *Journal of Labor Economics*, *8*(2), 202-229.

OECD (2014). PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014); PISA, OECD Publishing.

Patrinos, Harry A; Cristóbal Ridao-Cano and Chris Sakellariou. 2006. *Estimating the Returns to Education: Accounting for Heterogeneity in Ability*. Policy Research Working Paper; No. 4040. World Bank, Washington, DC.

Psacharopoulos, George. 1994. "Returns to Investment in Education: A Global Update." *World Development*, 22, 1325-1343.

Schoellman, T. (2012). "Education quality and development accounting." *The Review of Economic Studies, 79*, 388-417.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20,* 518-529.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews (Ed.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80 - 108). New York: Cambridge University Press.

Table 1. GDP and years of schooling for the Mexican states.

	2010 GDP	2016 GDP	2010 GDP per worker	2016 GDP per worker	2010 GDP per hour worked	2016 GDP per hour worked	2010 Years of schooling	2016 Years of schooling
Aguascalientes	152,205	216,703	401,990	471,061	164	190	10.0	10.5
Baja California	428,163	524,405	376,012	403,685	156	173	9.7	10.1
Baja California Sur	110,656	133,147	441,152	450,063	186	193	10.3	10.6
Campeche	753,969	600,771	2,767,822	1,836,844	1,104	780	9.4	10.0
Coahuila	489,952	583,873	541,627	546,131	220	232	10.2	10.5
Colima	81,992	101,336	340,332	364,120	139	158	9.7	10.1
Chiapas	270,989	288,692	207,848	178,923	89	81	7.3	7.7
Chihuahua	417,796	539,144	378,700	390,336	160	166	9.4	10.0
Ciudad de Mexico	2,446,910	2,974,071	716,881	843,660	292	358	11.3	11.7
Durango	169,268	202,998	363,232	341,422	150	148	9.5	9.9
Guanajuato	517,169	691,613	313,169	356,276	127	145	8.6	9.2
Guerrero	211,891	238,468	226,295	210,202	94	92	8.3	8.5
Hidalgo	206,304	264,242	253,440	261,985	106	113	9.0	9.1
Jalisco	925,372	1,161,406	377,290	412,538	156	185	9.5	10.1
México	1,226,814	1,478,587	238,680	241,531	93	100	9.6	10.0
Michoacán	329,767	406,185	260,443	261,831	110	123	8.2	8.5
Morelos	174,984	191,797	286,990	278,932	117	119	9.8	9.9
Nayarit	97,786	119,106	281,850	277,000	119	125	9.5	9.9
Nuevo León	1,025,184	1,228,744	627,538	642,637	252	272	10.5	10.9
Oaxaca	228,089	257,146	218,661	195,321	92	86	7.9	8.1
Puebla	469,968	557,877	270,890	254,118	111	114	8.7	9.0
Querétaro	287,403	385,622	464,254	581,925	195	241	9.9	10.3
Quintana Roo	195,149	262,760	386,601	393,597	150	161	9.7	10.2
San Luis Potosí	269,397	346,378	355,697	379,290	151	168	9.4	9.8
Sinaloa	312,655	381,109	366,133	385,254	150	172	10.0	10.4
Sonora	431,502	570,174	489,434	547,739	202	234	10.1	10.7
Tabasco	525,012	523,613	764,475	692,312	305	291	9.7	9.9
Tamaulipas	448,215	489,100	414,387	388,932	168	172	10.0	10.2
Tlaxcala	88,810	97,665	234,560	218,018	97	93	9.5	9.9
Veracruz	718,149	811,543	307,670	319,747	124	140	8.6	8.8
Yucatán	196,150	242,005	295,846	294,169	121	131	9.0	9.3
Zacatecas	144,731	157,898	363,048	324,869	153	141	8.9	9.4
All Mexico	14,352,401	17,028,177	395,809	400,283	162	172	9.4	9.8
Mean			447,905	429,515	183	184	9.4	9.8
Std. Dev.			437,608	294,048	174	124	0.8	0.8

* GDP is measured in millions of 2013 pesos. GDP per hour worked and per worker are in constant 2013 pesos.

Table 2. PISA Mathematics test scores for the Mexican states.

	PISA test scores 2003	**PISA test scores average 2003-2006-2009-2012**
Aguascalientes	639	634
Baja California	482	487
Baja California Sur	518	447
Campeche	364	370
Coahuila	541	544
Colima	555	544
Chiapas	242	279
Chihuahua	544	580
Ciudad de Mexico	577	653
Durango	586	527
Guanajuato	550	546
Guerrero	286	263
Hidalgo	556	537
Jalisco	637	618
México	499	564
Michoacán	538	506
Morelos	608	575
Nayarit	514	491
Nuevo León	703	704
Oaxaca	486	432
Puebla	518	526
Querétaro	538	605
Quintana Roo	451	458
San Luis Potosí	498	480
Sinaloa	491	513
Sonora	492	492
Tabasco	284	288
Tamaulipas	511	488
Tlaxcala	444	440
Veracruz	398	448
Yucatán	429	444
Zacatecas	520	519
Mean	**500**	**500**
Std. Dev.	**100**	**100**
Max-Min	**461**	**425**

Table 3. Adjustment of PISA test scores for interstate migration, autoselection of interstate migration, and international migration.

Variable	Obs	Mean	Std. Dev	Min	Max
Average 2003-2012 standardized	32	500	100	263	704
Average + interstate migrants	32	500	88	286	647
Average + interstate migrants + adjustment by educational category	32	496	78	296	602
Average + interstate migrants + adjustment by educational category + international migrants	32	497	78	298	602

Table 4. Elasticity of years of schooling with respect to quality of schooling, using 2010 Census data.

	OLS		Instruments	
	Elasticity	**Implied η**	**Elasticity**	**Implied η**
Coefficient	0.239***	0.194	0.247***	0.198
Std. Dev.	(.051)		(.063)	

*, **, and *** refer to significance of the coefficients at the 10%, 5% and 1% levels, respectively.

Table 5. Percentage of variability of income attributed to human capital, using 2010 Census data.

	Share Q quality		Years of schooling	Total sum of quality and quantity Q and S		Top and bottom 5 states Q and S		Top and bottom 3 states Q and S	
	ln GDP per worker	***ln GDP per hour***		***ln GDP per worker***	***ln GDP per hour***	***ln GDP per worker***	***ln GDP per hour***	***ln GDP per worker***	***ln GDP per hour***
Additive model									
Average 2003-2012 standardized	0.35	0.34	0.18	0.53	0.52	0.57	0.58	0.75	0.77
Average + interstate migrants	0.33	0.33	0.18	0.51	0.51	0.56	0.56	0.74	0.77
Average + interstate migrants + adjustment by educational category	0.31	0.31	0.18	0.49	0.49	0.54	0.55	0.72	0.75
Average + interstate migrants + adjustment by educational category + international migrants	0.31	0.31	0.18	0.49	0.49	0.54	0.55	0.73	0.75
Multiplicative model									
OLS				0.70	0.70	0.65	0.65	0.83	0.86
Instruments				0.70	0.70	0.65	0.65	0.83	0.86

Table 6. Percentage of variability of income attributed to human capital using 2016 ENOE.

Panel A. Results additive model

	Share Q quality	**Years of schooling**	**Total sum of quality and quantity Q and S**
Average 2003-2012 standardized	0.32	0.17	0.49
Average + interstate migrants	0.31	0.17	0.48
Average + interstate migrants + adjustment by educational category	0.29	0.17	0.46
Average + interstate migrants + adjustment by educational category + international migrants	n/a	0.17	n/a

Panel B. Results multiplicative model

	OLS		**Instruments**		
	Elasticity	η	Elasticity	η	F
Coefficient	0.178**	0.151	0.467**	0.318	3.196
Std. Dev	(0.077)		(0.209)		
Variation explained by human capital		0.41		0.39	

*, **, and *** refer to significance of the coefficients at the 10%, 5% and 1% levels, respectively.

Table 7. Results of multiplicative model for GDP per hour for years 2010 and 2016 using only migrants.

	2010					2016				
	OLS		**Instruments**			**OLS**		**Instruments**		
	Elasticity	η	Elasticity	η	F	Elasticity	η	Elasticity	η	F
Coefficient	0.301***	0.232	0.332***	0.249	10.104	0.203***	0.167	0.240***	0.193	10.056
Std. Dev	(0.050)		(0.074)			(0.045)		(0.067)		
Variation explained by human capital	0.65		0.65			0.55		0.55		

*, **, and *** refer to significance of the coefficients at the 10%, 5% and 1% levels, respectively.

Table 8. Variability of income attributed to human capital using PISA 2003

	Results for 2010- using PISA 2003			Results for 2016- using PISA 2003		
	Share Q quality	Years of schooling	Total sum of quality and quantity Q and S	Share Q quality	Years of schooling	Total sum of quality and quantity Q and S
Additive model						
PISA 2003 standardized	0.31	0.18	0.49	0.27	0.17	0.44
Average + interstate migrants	0.30	0.18	0.47	0.27	0.17	0.43
Average + interstate migrants + adjustment by educational category	0.28	0.18	0.46	0.25	0.17	0.42
Average + interstate migrants + adjustment by educational category + international migrants	0.28	0.18	0.46	n/a	0.17	n/a

Table 9. Variability of income attributed to human capital when including Campeche and Tabasco

	Baseline (30 states)		Baseline + Tabasco		Baseline + Tabasco + Campeche	
	2010	2016	2010	2016	2010	2016
Number of states	30	30	31	31	32	32
Var ln GDP per hour	0.09	0.12	0.10	0.13	0.23	0.21
Quantity of schooling	0.18	0.17	0.15	0.10	0.07	0.07
Additive model						
Quality of schooling	0.31	0.29	0.16	0.19	0.00	0.05
Total (quality + quantity)	0.49	0.46	0.31	0.29	0.07	0.12
Multiplicative model						
OLS	0.70	0.41	0.57	0.34	0.28	0.23

Appendix A: Methodology for adjusting test scores for migration between states, autoselection of migrants, and international migration.

We adjust the PISA test scores of each state following Hanushek et al. (2017). First, we adjust test scores for interstate migration and for the autoselection of interstate migration, and then we adjust test scores for international migration and its autoselection.

<u>Interstate Migration</u>

In our base model, we assign each individual the PISA test score of their state of residence. To correct test scores by interstate migration, we distinguish between an individual's state of birth and their state of residence. If an individual resides in a state other than their birth state, we assume the individual went to school in their birth state and therefore assign the birth-state PISA test scores to the individual. First, for each state, we group residents according to their birth state. For instance, in Aguascalientes 70.8% of the were born there, while 7% were born in Mexico City, 6% in Zacatecas, 5% in Jalisco and so on for each of the 32 states. We also form a category of state residents who were born outside Mexico (international migrants). To adjust the state average test score for interstate migrants in the case of Aguascalientes, we would then multiply the PISA test score of Aguascalientes by 70.8%, and to this add 7% of the score of Mexico City, and so on. In the case of international migrants, we assign them the average score of their state of residence initially. We correct for international migration as the last step of these adjustments.

Table 1A shows the 2003-2012 average Mathematics PISA test scores by state, standardized with mean 500 and standard deviation 100. After correcting for interstate migration (columns 3 and 4) the average score is still 500 while the standard deviation falls to 88.

Correction for migrant autoselection bias

To correct for migrant interstate autoselection, we separate workers into two groups, those with up to 12 years of schooling and those with 13 years or more, with the objective of identifying individuals with access to higher education. For instance, 70.8% of the residents of Aguascalientes were born in the state, this group can be split into 55.8% which have up to 12 years of schooling and 15% with 13 years or more. Then, for each state we subdivide the individual PISA test scores according to whether at least one of the test taker's parents has some higher education. We then make the assumption, as in Hanushek et al. (2017), that we can assign individuals with higher education the PISA test score of children whose parents have higher education, and vice versa for individuals without higher education. We then adjust PISA test scores by weighing them according to interstate migration, but adjusting separately for residents with higher education and those without. As a result of this adjustment, the average 2010 PISA score falls to 496 and the standard deviation falls to 78 (column 5 in Table 1A).

Correction for international migration.

Our sample (10% of the Population Census) contains 9613 working foreigners from 92 countries, out of 3,304,715 total workers, hence less than 0.3% of the working population are international immigrants. To obtain test scores for these migrants, we use OECD (2004, Table 2.3c) PISA mathematics test scores, where we take the mean, standard deviation, and 75 and 90 percentiles approximating the methodology of Hanushek et al. (2017). For the countries for which we do not have PISA scores, we approximate the scores using countries that are similar or geographically close.[14]

To adjust for the selectivity of international migration in Mexico we follow Hanushek et al. (2017) who show that in the case of the U.S. such selectivity is significant. We start by computing the selectivity parameter p for each country, which indicates the percentile of the home country distribution from which the average immigrant comes, from educational degrees primary (pri), secondary (sec) or tertiary (ter). The equation that Hanushek et al. (2017) use to calculate the selectivity parameter is the following:

$$p = s_{MX}^{pri} * \frac{1}{2} s_{\mathrm{home}}^{pri} + s_{MX}^{\sec} * \left(s_{\mathrm{home}}^{pri} + \frac{1}{2} s_{\mathrm{home}}^{\sec} \right) + s_{MX}^{ter} * \left(s_{\mathrm{home}}^{pri} + s_{\mathrm{home}}^{\sec} + \frac{1}{2} s_{\mathrm{home}}^{ter} \right)$$

Where s_{MX}^{pri} would indicate the proportion of the migrants from a particular home country working in Mexico who only have primary education, and s_{home}^{pri} would indicate the proportion of the population of the home country with only primary education. For instance, if from a country where schooling is (0.1, 0.1, 0.8)- indicating 10% have primary education, 10% have secondary education and 80% tertiary education, we have that immigrants into Mexico only have primary education, then p=0.05. If from a country with low education (0.8, 0.1, 0.1) all of those who reside in Mexico have tertiary education, p=0.95. If from a country with equal proportions of educational degrees (0.33, 0.33, 0.33) the workers in Mexico have the same proportions then we would have

[14] For North Korea we use South Korea, and for Macao and Taiwan we use China. For other African and Asian countries we use the scores from Tunisia which is the only country available. For Center and South America we group the data according to the three countries for which we have PISA scores: Brazil, Mexico and Uruguay. For the rest of Europe we use Greece. According to the 2015 PISA test, which was administered in more countries, we use Germany for the case of England and we use Greece for the case of Israel.

p=0.5. The proportions of immigrants in Mexico with different educational degrees we obtain directly from the data, and the proportions with the respective degrees in the home countries we obtain from the database in Docquier, Lowell and Marfouk (2009, http://www.rnim.org/uploads/1/6/3/4/16347570/dm_dataset.xls). We find countries that are geographically close, such as USA and Guatemala have p of 0.4 and 0.52 while countries that are farther away have higher values, such as Japan and Ecuador with values of 0.8. As in Hanushek et al. (2017) we then adjust PISA test scores given the value of p as follows:

$$scoreselp_j = invnormal(p_j)*pisa\_sd_j + pisa\_av_j$$

where the invnormal function is the inverse of the normal, $pisa\_sd_j$ is the standard deviation of the mathematics PISA scores for country j, and $pisa\_av_j$ is the average score for country j[15]. The last two columns of Table 1A show the test scores corrected for international migration.

[15] We also compute the values for the 75 and 90 percentiles, $scoresel75_j$ = invnormal (.75)*$pisasd_j$ + $pisaav_j$; and $scoresel90_j$ = invnormal (.90)* $pisasd_j$ + pisaav; and the correlation coefficients between the estimated values of scoresel75, scoresel90, and the real is of 0.98 for the 75th percentile and 0.97 for the 90th percentile.

Table 1A. PISA test scores by state, adjustment of test scores for migration between states, autoselection of migrants and international migration.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
State	PISA Average 2003-2012	+ With interstate migration.		+ Adjusted for selective migration		+ Adjusted for international migrants	
		2010	2016	2010	2016	2010	2016
Aguascalientes	634	615	618	600	604	600	604
Baja California	487	500	494	487	481	489	482
Baja California Sur	447	473	456	480	456	484	457
Campeche	370	380	380	384	384	386	385
Coahuila	544	546	543	544	539	544	539
Colima	544	543	542	536	534	536	534
Chiapas	279	288	286	296	293	298	295
Chihuahua	580	564	565	545	546	545	546
Cd. de México	653	617	620	589	593	589	593
Durango	527	530	529	531	528	531	528
Guanajuato	546	549	548	551	550	551	550
Guerrero	263	286	279	305	294	306	295
Hidalgo	537	543	541	544	541	544	541
Jalisco	618	601	605	585	592	585	592
México	564	573	575	564	567	564	567
Michoacán	506	509	508	511	507	511	508
Morelos	575	538	544	513	523	513	523
Nayarit	491	500	499	502	501	502	501
Nuevo León	704	647	654	602	613	602	613
Oaxaca	432	436	434	439	437	440	437
Puebla	526	524	524	522	524	523	524
Querétaro	605	595	592	593	590	594	591
Quintana Roo	458	440	442	421	417	422	417
San Luis Potosí	480	491	487	499	495	501	496
Sinaloa	513	513	512	512	510	512	510
Sonora	492	497	497	497	497	497	497
Tabasco	288	312	304	335	322	336	322
Tamaulipas	488	492	490	487	483	487	484
Tlaxcala	440	462	457	476	470	476	470
Veracruz	448	453	453	456	456	457	457
Yucatán	444	447	445	451	446	452	447
Zacatecas	519	525	524	530	528	530	528
Mean	500	500	498	496	494	497	495
Standard deviation	100	88	90	78	82	78	81

Table 1B. Estimates of rates of return to schooling by state

State	**Rate of Return**
Aguascalientes	0.112
Baja California	0.120
Baja California Sur	0.087
Campeche	0.105
Coahuila	0.130
Colima	0.089
Chiapas	0.087
Chihuahua	0.117
Cd. de México	0.140
Durango	0.094
Guanajuato	0.089
Guerrero	0.074
Hidalgo	0.094
Jalisco	0.098
México	0.091
Michoacán	0.078
Morelos	0.095
Nayarit	0.075
Nuevo León	0.156
Oaxaca	0.082
Puebla	0.086
Querétaro	0.103
Quintana Roo	0.093
San Luis Potosí	0.094
Sinaloa	0.089
Sonora	0.102
Tabasco	0.094
Tamaulipas	0.122
Tlaxcala	0.091
Veracruz	0.097
Yucatán	0.087
Zacatecas	0.083
Mean	0.098
Variance	0.0003

Table 2B. 2010 Human capital estimates using the multiplicative model.

	OLS	Instruments
Aguascalientes	5.24	5.14
Baja California	5.33	5.24
Baja California Sur	5.03	4.93
Coahuila	5.47	5.37
Colima	5.01	4.92
Chiapas	4.72	4.63
Chihuahua	5.27	5.17
Cd. de México	5.64	5.55
Durango	5.03	4.93
Guanajuato	4.90	4.80
Guerrero	4.71	4.62
Hidalgo	4.96	4.87
Jalisco	5.09	4.99
México	5.03	4.94
Michoacán	4.73	4.63
Morelos	5.09	4.99
Nayarit	4.82	4.72
Nuevo León	5.68	5.58
Oaxaca	4.73	4.63
Puebla	4.87	4.78
Querétaro	5.18	5.08
Quintana Roo	5.07	4.97
San Luis Potosí	5.04	4.95
Sinaloa	5.07	4.97
Sonora	5.17	5.08
Tamaulipas	5.36	5.26
Tlaxcala	5.00	4.90
Veracruz	4.98	4.89
Yucatán	4.90	4.80
Zacatecas	4.83	4.74
Mean	5.06	4.97
Std. Dev.	0.25	0.25

Table 3B. Estimates of Human Capital using the Additive model.

	Standardized difference in PISA test (Quality Measure)	Human Capital (h) Additive model
Aguascalientes	3.81	1.46
Baja California	2.41	1.20
Baja California Sur	2.35	1.23
Campeche	1.10	0.95
Coahuila	3.11	1.35
Colima	3.00	1.29
Chiapas	0.00	0.59
Chihuahua	3.12	1.30
Cd. de México	3.67	1.54
Durango	2.94	1.27
Guanajuato	3.20	1.24
Guerrero	0.09	0.69
Hidalgo	3.10	1.25
Jalisco	3.62	1.38
México	3.35	1.35
Michoacán	2.68	1.12
Morelos	2.71	1.25
Nayarit	2.58	1.21
Nuevo León	3.84	1.50
Oaxaca	1.78	0.94
Puebla	2.83	1.18
Querétaro	3.73	1.43
Quintana Roo	1.59	1.05
San Luis Potosí	2.54	1.19
Sinaloa	2.70	1.27
Sonora	2.51	1.24
Tabasco	0.47	0.87
Tamaulipas	2.39	1.22
Tlaxcala	2.24	1.15
Veracruz	2.00	1.04
Yucatán	1.94	1.06
Zacatecas	2.92	1.22
Mean	2.51	1.19
Variance	1.00	0.04