Labour market, reading skills and gender

Large scale assessment of reading literacy for Chile

Dante Contreras - Esteban Puentes - Valentina Riveri

Abstract

Given the relevance of human capital in a knowledge-based economy to policy, it is essential to improve how it is measured, considering not only quantity, but also quality. We take advantage of a new survey with several tests that approximate several quality characteristics of human capital. We focus on a reading comprehension test especially designing to measure the reading and comprehension ability of Chileans. The survey also includes information about cognitive abilities and psychological traits. By estimating a simple model of wage determination and labor participation, we find that the age, education, and marital status affect the odds of participating in the labor market. Additionally individuals with better reading comprehension skills tend to receive higher wages, confirming the hypothesis that they intrinsically have higher productivity levels. Despite this, cognitive abilities and psychological traits are not related to labor participation or wage determination.

1. Motivation and Literature review

Human capital was introduced as a concept in the early 1800,[1] but it was not until the 1960s that its theory was proposed and developed. With clarity on human capital's impact on growth (Becker, 1964), many researchers have sought to distinguish determinants that lead to greater productivity in some sort of economic context.

One of the main problems that empirical literature has faced is measurement. This is, definitive consensus on how to measure human capital does not exist. In general, indirect indicators are used such as school enrollment rates (Barro, 2001); measurements that do not consider quality and only capture certain aspects of human capital.

Several researchers, concerned with finding a better measurement, have used scores obtained from international assessments, some in science and mathematics (Hanushek & Kimko, 2000), others in reading (Coulombe & Tremblay, 2006), and some more adventurous (Altinok, 2007). These new indicators are based on a pool of international surveys concerning student assessment (TIMMS, PIRLS, PISA, SACMEQ, PASEC, LLCE and MLA).

Hanushek and Kimko (2000) work with direct measures of cognitive skill comparing them across countries. Their aim is to illustrate how the inclusion of quality measures can explain cross-country growth, finding that quality measures replace quantity in demonstrating the importance of human capital (Labour-force quality) for national growth.

Coulombe and Tremblay (2006), working with the results of the 1994-1998 International Adult Literacy Survey (IALS), derive time-series data comparable at a cross-country level. They find that synthetic measures of human capital contain more information on relative growth than the school data does, indicating that literacy test scores are more direct and accurate measures of human capital than years of schooling.

The above studies have confirmed the importance of measuring multiple dimensions of human capital and provide a framework for our empirical analysis.

[1] Adam Smith, H Von Thünen, and Irving Fisher.

1.1 Literacy as quality measurement of human capital

Several investigations have used literacy as a measure of quality of human capital, for instance Johnston (2004), and Green and Riddell (2001). This is due to access to international tests measuring literacy and allow for comparability among countries with a single measure of human capital.

The development of direct measures of skill attainment, such as the IALS, offers labour economists a powerful new tool to help explain market outcomes. For instance, Johnston (2004) used the IALS data to investigate the relationship among educational attainment, literacy, and economic growth. The study suggests that differences in average skill levels account for 55% of the differences in economic growth during the period 1960-1994 among OECD countries.

Another line of research has focused primarily on individuals, analyzing the relationship between literacy and labour market variables. Green and Riddell (2001) find that the impact of education on earnings is comprised of two separate effects. The first is an increase in revenue associated with the development of specific knowledge and skills through increased education. The second effect is an increase in income resulting from stronger literacy skills, which are also associated with higher education levels (Green & Riddel, 2001).

1.2 Evidence for Chile: Labour market and literacy

Different national studies have shown that economic growth depends largely on the stock of human capital. In this line, Gallego and Loayza (2002) estimate that the GDP growth in Chile during 1961-1985 was 2.54%, where the labour force (human capital) was the main driver (1.45%), followed by physical capital (0.95%) and, finally, by total factor productivity (0.14%). Meanwhile, for 1986-2000, the economy grew at an average rate of 6.64%, where physical capital was the main driver (2.47%), followed by the labour force (2.30%), and finally by total factor productivity (1.87%) (Gallego & Loayza, 2002). This leads to the conclusion that human capital has decreased its contribution to GDP growth.

Several researchers have estimated the returns to human capital and education using Mincerian equations for Chile. These studies have reported an increase in the variance of wages for individuals with equal years of schooling (Mizala & Romaguera, 2003) (Contreras & Gallegos, 2011).

There is empirical evidence (Contreras, Melo, Ojeda, 2005) on the effect of unobservable omitted variables on the returns to education (wages). Skill and quality of education are most often mentioned. A specific example of a study that includes these omitted variables is Bravo et al. (2002). They analyze the relationship between literacy and labour market variables, specifically job opportunities and incomes[2] (Bravo, Contreras, & Larrañaga, 2002) and find that greater literacy skills are associated with higher earnings, even controlling for years of schooling and other earnings related variables.

1.3 Key Questions

The present study addresses the relationship between reading skills and labor market variables in Chile. Although Chile improved its performance in the literacy PISA test between 2000 and 2009, matching and even surpassing the results of countries with similar GDP, it is still necessary to improve the results in those most vulnerable socioeconomic groups, and set new goals consistent with the international context.

[2]Women were not considered in order to isolate the labor market participation decisions.

While Chilean PISA literacy scores increase with income, reaching and even exceeding the OECD average (493) for the wealthiest group (1.5 average or higher), the lowest income group scores far below this goal (SIMCE – Unidad de Curriculum y Evaluación Ministerio de Educación de Chile, 2010) Table 1 literacy scores and average income.

At the same time, compared with OECD countries, Chile is under the trend, i.e. the GDP per capita has presented scores lower than expected (Illustration 1 GDP per capita and PISA reading scores for OECD countries, 2009.). This prompts promotion of quality policies designed to continue development.

This study seeks to promote discussion based on two key questions:
Is there a direct wage effect from literacy skills?

Is the impact differentiated by gender?

We work with labor market variables and results of the reading comprehension test in 2011, conducted by the National Council for Culture and the Arts.

2. Methodology

We follow much of the previous literature in our formal analysis, using a Heckman two-step, or correction, method. For instance, women with high reservation wages may not want to participate in the labour market and, if this selection is due to non-observable factors, we could be missing cases non-randomly; therefore our results may not be applicable to the entire population. In order to avoid this selection problem, we use Heckman methodology, modeling the wage equation jointly with the selection into the labour market. We also use OLS when we want to estimate the effect of some factors on a continuous dependent variable, in this case logarithm of hourly wages.

We use the below wage equation

Where is the log of hourly wage, observed variables relating to the i'th person's productivity, and is an error term. W is observed only for workers, i.e. only people who receive a wage.

Then, the sample selection (1); i.e. being in the labour force, so is observed

Where is a latent variable that captures the propensity to work, observed variables for the i'th person's, and is the stochastic error representing all the unobservable variables that may influence the propensity to work. Even though is not observed, can define an indicator variable using wage observability. Then, wage will be observed if >>0,

Heckman methodology approaches the problem as an omitted variable, hence estimating it will solve the selection bias problem. The omitted variable would correspond to the propensity to work.

(2) Where is the correlation coefficient between the error terms of both equations (selection and wage), is the standard deviation of the error in the wage function, and is the inverse of mills ratio.

The selection equation and earnings equation are jointly estimated in a Heckman procedure. Equation 1 is the estimation of the parameters by the probit method using the full sample (workers and non-workers) that estimates the probability of a person working. These estimates are used to calculate the value of for each individual in the sample. Both are jointly estimated by full maximum likelihood using the Heckman procedure in STATA.

3. Data

In 2011, the National Council of Arts and Culture, the governmental entity responsible for implementing public policies for the development of culture in Chile, commissioned the first study of reading habits and comprehension nationwide. This study considers two sections, first a socioeconomic survey and the other a reading comprehension test. The latter assessed the literacy proficiency of representative samples of individuals aged between 15 and 65. This indicator provides a direct measure of human capital quality.

The socioeconomic survey collected information on non-cognitive skills and individual's self-perception, which allow us to control for self-efficacy, autonomy, and locus control skills that can impact the probability of participating in the labour market and its productivity. If these types of factors are not considered, there is the risk of overestimating the impact of literacy improvements.

The following describes the literacy test and presents relevant descriptive statistics.

3.1 Reading comprehension test

One of the objectives of this measurement was to approximate actual reading; therefore, the texts in the assessment tool were taken from day-to-day experiences of the groups tested. Comprehension was assessed through multiple choice and constructed response questions. Additionally, questions from the reading assessment questionnaire Pisa 2009 were included for testers older than 15. This was so the study could be compared internationally.

3.1.1 Performance levels

The results of the Reading Comprehension Test are centered on a scale with a mean of 100 and a standard deviation of 20. The reliability coefficient of the set of items corresponds to 0.84, and the methodology of Item Response Theory (IRT) was used to generate scores.

The scores have been nested in five performance levels (1 through 5), which are determined by skill levels in each of the following tasks:

1. Extracting and locating information
2. Interpreting a text
3. Reflecting on the form and content of a text

3.2 Descriptive Statistics

Below is a statistical description of some socioeconomic variables, comparing them with the National Socioeconomic Characterization Survey (CASEN), conducted by the Ministry of Social Development every two or three years since 1987. This also includes descriptive statistics of the reading comprehension test scores.

3.2.1 Socioeconomic variables (*Table2 Descriptive* Statistics*)*

We used data for the population over 15 years, 1217 people. Of these, 288 did not answer the reading comprehension test, 9 did not answer the survey, and one person had inconsistency responses concerning to labour status. Thus, our final sample consists of 928 observations. Estimations consider three groups of people; the entire population, those over 21, and people who are not currently in school.

From the final sample, 40% have high school education or less (CASEN 2009 45%); while 11.6% has higher education (14.8% CASEN 2009). Our sample has, on average, 11.3 years of education.

In relation to marital status, 41% are single, more than 33% are married, about 14% live with or have a partner (CASEN 2009 14.49%) and 7.6% of interviewees reported being separated (CASEN 2009 6.14%). Considering these statistics, a marital status dummy was created, which shows that 47% of the sample is married or living with a partner.

Of those interviewed, 60.7% reported participating in the labour market (2009 CASEN 61%). At the time of the survey, 57% were employed (CASEN 2009 55.78%) and 2.8% were unemployed (5.34% CASEN 2009). Looking at the gender distribution, 79% of men were participating (CASEN 2009 79%) while women's participation was less than 47% (CASEN 2009 46%).Overall, our sample has a 57% employment rate.

Finally, from the sample 46% live in the Metropolitan Region, they have an average age of 38.8 years old, and 7.1 years of work experience in their current job.

3.2.2 Reading comprehension test

Below is the descriptive statistics for the reading comprehension test, specifically the average across different population groups. The descriptive statistics are delivered with rescaled scores, and then the estimation uses the standardized scores, which have a mean of 0 and standard deviation of 1.

The results of the reading comprehension test results are given as five levels. Level one consists of locating one or more pieces of information explicitly and discriminating between little or no information that compete with each other, recognizing the main topic in texts with familiar content, and making simple connections between content and everyday life. People who reach level two are able locate one or more pieces of information with multiple criteria in texts of different purposes, identify the main idea, perform simple relationships, and successfully discriminate information. Additionally, they construct meaning within despite being given unimportant information and can make low-level inferences and comparisons between the text and external knowledge. Level three requires recognizing the relationship among dispersed fragments of the text information and discriminating between competing ideas. Readers at this level also integrate various parts of the text to identify a main idea and to understand a relationship or construe the meaning of a word or phrase. They create relationships, provide explanations and evaluate the text based on daily or in some cases more specifics knowledge.

People that reach the fourth and fifth levels are able to locate and organize information in texts with different purposes, deducting what is relevant. They also use high-level inferences based on the text to interpret a sections within the text or to explain features. They handle ambiguity and different ideas generated by the text and critically assess or formulate hypotheses derived from expertise in relation to the text's subject. Additionally, they demonstrate an adequate understanding of long and complex texts with unfamiliar content.

When analyzing the distribution by level as shown in Table. 3, we can observe only 16% of the population reaches level four or five (*Table 3 Distribution of performance levels).*

Unsurprisingly, higher educational levels are associated with higher scores. People with *Eight years of education or less* have significantly lower scores. People with at *least undergrad education* have significantly higher scores, as shown in Table 4. Meanwhile there are not significant differences by gender (*Table 4 Distribution Educational level and gender.)*

4. Results

In this section we present the results for 928 people, a sample representative of the Chilean population over 15 years of age. The estimates were made for three samples of the population: the whole population, those older than 21, and for those who were not in school.

The models analyzed correspond to a basic one that includes classical variables to explain income and standardized test scores as a measure of human capital in the selection equation. The basic model was also estimated including non-cognitive variables, math skill variables, and both simultaneously for the whole population.

4.1 Basic model

The basic model includes variables usually presented in the international literature; these results are reported in Table 5 *Basic model and model including standardized test scores of reading comprehension*. The baseline specification includes experience, gender, years of schooling, a dummy for living in the metropolitan region, age, and the standardized reading test scores. The selection equation includes gender, years of schooling, a dummy for living in the metropolitan region, age, marital status, and number of children older than 2 and younger than 15 who live at home.

The experience variable corresponds to the number of years at the current job. Experience squared was also included. For the hourly earning´s equation, the coefficients of each variable are precisely estimated and have the expected sign. We include a polynomial for potential experience to capture the non-linearities with earnings, but the coefficients show a similar result to the concave profile found in the literature.

Results show significant differences in wages between regions. For instance, people living in the metropolitan region have 11% higher average hourly wages than the rest of the country. Results also indicate different wages between genders, with men having 31% higher hourly wages.

The relationship between years of schooling and hourly earnings follows national and international evidence. For Chile, as indicated by Contreras and Melo (Contreras; Melo; Ojeda, 2005), and Mizala and Romaguera (Mizala, Alejandra; Romaguera, Pilar, 2001) the return to education is around 9-14% on hourly earnings. Years in the current job also impact hourly earnings, where each additional year increases wages by 3%.

Years of education´s coefficient is positive and statistically different from zero at the 10% level, and its magnitude indicates that an additional year of education increase hourly wages by 10%.

Interestingly test scores capture an additional source of variance not captured by years of schooling. The latter is a standardized variable so its interpretation should be in standard deviation units, i.e. each additional standard deviation increases hourly wages by 6.7% This shows the importance of reading comprehension in the labour market, indicating that individuals with higher reading comprehension levels tend to receive higher wages, confirming the hypothesis that they have intrinsically higher productivity levels.

This result is consistent with Bravo, Contreras and Larrañaga (Bravo, David; Contreras, Dante; Larrañaga, Osvaldo, 2002) for Chile that shows that a higher reading skill levels are associated with higher income. With regard to international studies, the results are consistent with Green and Riddell (Green & Riddell, 2007) who found that in terms of income generation, reading skills play an important role. Additionally, it can be argued that some of the observed wage inequality in Chile is due to reading comprehension differences; therefore improving this variable in the most vulnerable sectors of the population may have effects not only on wages, but also in their distribution.

Years of schooling also positively impact the employment probability model; an additional year of education increases participation by 5.4%.

The probability of working differs by gender in the selection equation, where men are 8.7% more likely to be in the labor market. The number of children in the household (2 < age <= 15) is statistically different from zero, decreasing being in the labor market by 17%. Number of babies in the household (age<=2 years) was also included. Babies impact in the likelihood of participating by 40%. Age is statically different from zero, increasing the probability of working by 23%. Living in the metropolitan region and marital status show no significant difference in the probability of participating.

4.2 Basic model including reading comprehension standardized test scores in the selection equation

In this specification the significance and magnitude of the coefficients behavior is stable in relation to the above model. It is noteworthy that test scores are not statistically different from zero in the selection model, i.e., not having a relationship with labor market participation. Indicating that the reading comprehension skills would not affect labor market entry, but would determine the worker's productivity. These results are reported in Table 6 Basic model.

4.3 Basic model for population over 21 years.

The following section provides the second specification estimated for population older than 21. These results are reported in Table 7. We can observe that the results do not vary significantly from the model estimated for the entire population. Years-of-schooling positively impact hourly earnings by 10%. Living on the metropolitan region affect hourly earnings by 12.9%.

The standardized scores stay significant in hourly earnings, where an additional standard deviation increases wages by 6.3%. One difference compared to the previous model is that age significantly impacts hourly wages, increasing them 0.6%. The impact magnitude of gender decreases, and men earn 21% more hourly.

The selection model results are also similar. Age affects the probability of labour market participation, but its magnitude is decreased, i.e. an additional year increased the likelihood of participation by 12%. Years of education also maintains its significant, but its magnitude also decreases, increasing the likelihood of participation by 4.8%. Consistent with the previous literature, the number of children younger than 2 years old decreases the probability of participating in the labour market by 27% (lower than previous models). Unlike previous models the number of children between 3 and 15 years old is not significantly different from zero.

4.4 Basic model for population not currently in school.

In this specification many variable maintains it significance and magnitude in comparison with the latter model, with exception of the standardized score, living in the metropolitan region, and the significance of babies and children in the household. These results are reported in Table 7.

The standardized score increase its magnitude; an additional standard deviation increases hourly earnings by 7%.

Meanwhile the number of children has similar effects as in the first model. In the employment selection equation, the number of children in the household ($2 < age <= 15$) is statistically different from zero, decreasing the likelihood to be in the labor market by 15%. The number of babies in the household (age<=2 years), however, does not significantly differ from zero.

Regarding living in the metropolitan area, the results resemble the first models, no significant difference in labor market participation.

4.5 Basic model was estimate including non-cognitive variables, math skills variables and both simultaneously for the whole population.

When analyzing the results by including non-cognitive and math skills, the results do not differ significantly; the skills variables do not impact the hourly wage or in the likelihood of labor market participation. Interestingly, when both skills are estimated simultaneously, years of schooling has similar results for the wage equation, but loses significance for the selection one. The standardized scores also lost their significance in the wage equation, perhaps indicating that these skills are related or proxies for each other.

These results are reported in Tables 8 and 9.

5. Conclusion and final remarks

This paper uses the reading comprehension study to analyze the relationship between literacy and labor market variables. Improving the analysis by considering both gender and a measure of human capital quality, the study underscores the importance of including quality measures along with the years of schooling, to better explain income variance.

For the hourly earnings equation, the main variables explaining hourly wages are experience, which has a nonlinear positive relation; region, with higher wages in the metropolitan area; gender as men tend to earn 30% more; and finally years of schooling, indicating that an additional year of education increases hourly earnings by 10%.

In the selection equation of employment, the probability of working differs by gender, 22% more for men. Meanwhile, years of schooling, age, and living in the metropolitan region increase it. With regard to the number of children and babies in the household, the first increase the likelihood of being in the labor market, but the number of children decreases it.

Finally, an additional standard deviation of the test increases hourly earnings by 6.4%. This shows the importance of reading comprehension in the labour market, indicating that individuals with higher levels of reading comprehension tend receive higher wages, confirming the hypothesis that they have intrinsically higher productivity levels. Chile is improving it population's reading skills and the test is negatively correlated with age, indicating a skill increase. However Chile still lags OECD countries. Based on this, and considering that the paper finds that reading is related to wages, it is even more important to solve this problem and constant improve the population's reading skills.

6. References

References

Altinok, N. (2007). Human Capital Quality and Economic Growth. *Institute for Research in the Sociology and Economics of Education* .

Barro, R. (2001). Education and economic growth. In OCDE, *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being* (pp. 14-41). París: OCDE.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. In G. S. Becker, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (pp. 45-144). Chicago: The University of Chicago Press.

Bravo, D., Contreras, D., & Larrañaga, O. (2002). *Functional Literacy and Job Opportunities.* Santiago: Departamento de Economía, Facultad de Ciencias Económicas y Administrativas, Administrativas.

Bravo, David; Contreras, Dante; Larrañaga, Osvaldo. (2002). *Functional Literacy and Job Opportunities.* Santiago: Documento de trabajo Nº195 Departamento de Economía Universidad de Chile.

Contreras, D., & Gallegos, S. (2011). Wage inequality in Latin America: a decade of changes. *Cepal Review 103* , 27-44.

Contreras, Dante; Melo, Emerson; Ojeda, Susana. (2005). ¿Estimando el retorno a la educación o a los no observables? *Estudios de Economía. Vol. 32* , 187-199.

Coulombe, S., & Tremblay, J.-F. (2006). Literacy and Growth. *Topics in Macroeconomics Volume 6, Issue 2* , Article 4.

Gallego, F., & Loayza, N. (2002). The Golden Period for Growth in Chile. Explanations and Forecasts. *Journal Economía Chilena (The Chilean Economy), Central Bank of Chile, vol. 5* , 37-67.

Green, D. A., & Riddell, W. C. (2007). *The Generation of Literacy and Its Impact on Earnings for Native Born Canadians.* Statistics Canada and Human Resources Development Canada, No.18.

Green, D., & Riddel, C. (2001). *Literacy, Numeracy and Labour Market Outcomes in Canada.* 89-552: Statistics Canada Catalogue number MIE2001008.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review, American Economic Association, vol. 90* , 1184-1208.

Lucas, R. E. (1988). Development, On the Mechanics of Economic. *Journal of Monetary Economics 22* , 3-42.

Mincer, J. A. (1974). Schooling, Experience, and Earnings. In J. A. Mincer, *Schooling, Experience, and Earnings* (pp. 0-167). Columbia University Press.

Mizala, A., & Romaguera, P. (2003). Remuneraciones y tasas de retorno de los profesionales chilenos. In P. Meller, & J. Brunner, *Oferta y Demanda de profesionales y técnicos en Chile. El Rol de la Información Pública.* Santiago: RIL editores.

Mizala, Alejandra; Romaguera, Pilar. (2001). *La legislación laboral y el mercado del trabajo en Chile: 1975-2000.* Santiago: Documentos de Trabajo 114, Centro de Economía Aplicada, Universidad de Chile.

Mizala, Alejandra; Romaguera, Pilar. (2003). Remuneraciones y tasas de retorno de los profesionales chilenos. *Documentos de Trabajo 114, Centro de Economía Aplicada, Universidad de Chile* .

Perticara, M. (2009). Brechas salariales por género en Chile: un nuevo enfoque. *CEPAL 99* , 133-149.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *The Journal of Political Economy, Vol. 94* , 1002-1037.

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review, Vol. 51* , 1-17.

SIMCE – Unidad de Curriculum y Evaluación Ministerio de Educación de Chile. (2010). *Resumen de Resultados PISA 2009 Chile.* Santiago: SIMCE.

Appendix

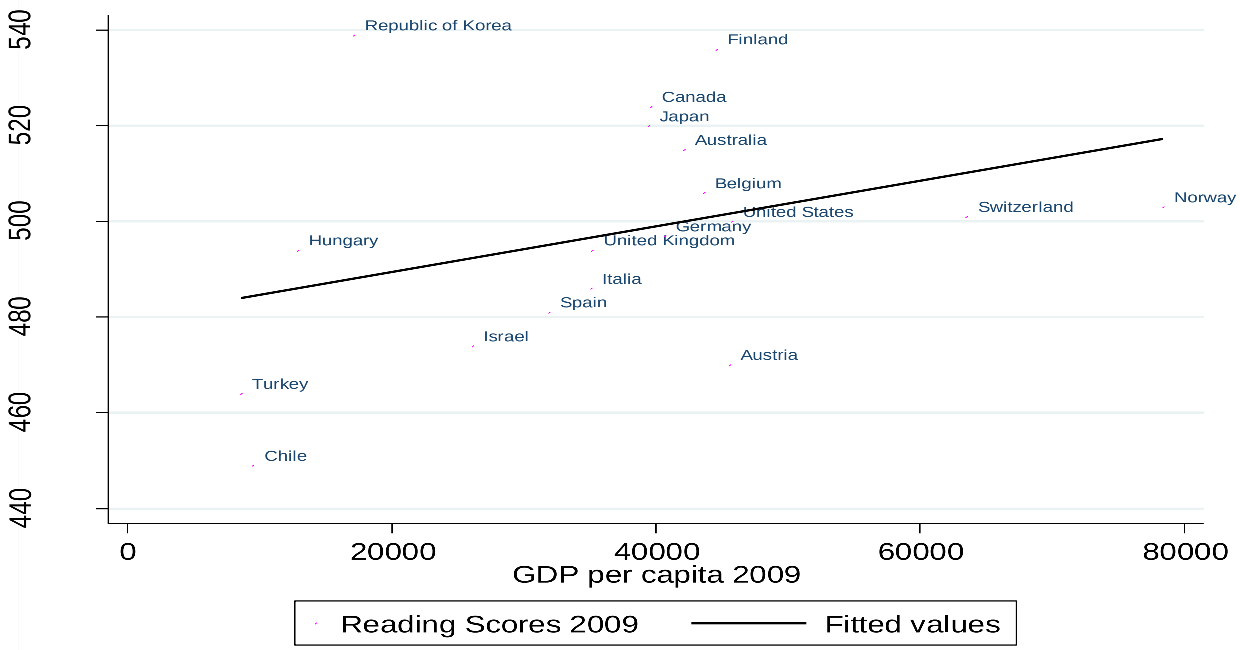
1 literacy scores and average income

Table 1 literacy scores and average income

Categories	Total Household Monthly Income	Average score Reading	Standard deviation	Cases post expansion factor	Sample cases	Establishments Sample Cases
1	less of <0.5 average>	418,57	3,85	69001	1524	167
2	<0.5 average or higher–less of 0.75 average>	440,62	4,09	52137	1153	166
3	<0.75 average or higher–less of average>	457,22	4,56	26269	591	152
4	<average or higher–less of 1.25 average>	466,04	4,09	18913	439	142
5	<1.25 average or higher–less of 1.5 average>	476,61	5,11	15966	375	122
6	1.5 average or higher	513,71	3,66	38606	1003	116
7	N/A	458,01	11,44	22103	486	115

2 GDP per capita and PISA reading scores for OECD countries, 2009

Illustration 1 GDP per capita and PISA reading scores for OECD countries, 2009.



3 Descriptive Statistics

Table2 Descriptive Statistics

Variables	Total sample Mean	Std. Dev	N	Male Mean	Std. Dev	N	Female Mean	Std. Dev	N	Description
Years of education	11.38	3.58	918	11.55	3.29	379	11.25	3.77	539	Approved educational years
Marital Status	0.47	0.49	928	0.44	0.49	381	0.50	0.50	547	1 Married, Live together or have a partner, 0 otherwise.
Employment situation	0.57	0.49	928	0.79	0.44	381	0.47	0.49	547	1 Employed; 0 otherwise.
Metropolitan Region	0.46	0.49	928	0.49	0.50	381	0.45	0.49	547	1 lives in the metropolitan region; 0 otherwise.
Children	0.60	0.82	928	0.54	0.78	381	0.64	0.85	547	Number of children less than 15 years old.
Age	38.81	15.41	928	35.97	15.11	381	39.10	15.49	547	Age in years
Experience	7.18	8.90	536	7.26	9.18	278	7.09	8.69	258	years of work experience in the current job

Source: Study of reading habits and comprehension 2011

4 Distribution of performance levels

Table 3 Distribution of performance levels

Level	Average Scores	Percentage
1	56	20,60%
2	82	38,20%
3	103	25,30%
4	121	12,70%
5	139	3,30%

5 Distribution Educational level and gender

Table 4 Distribution Educational level and gender

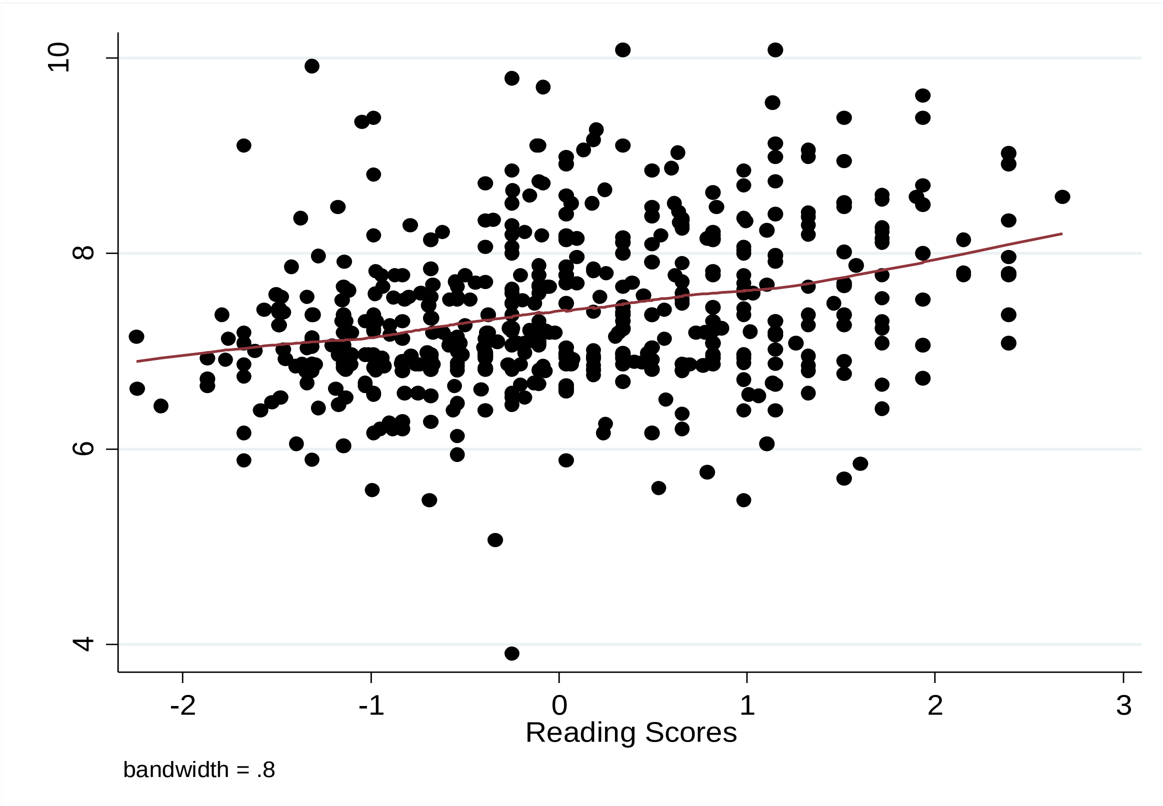
	Total sample			Male			Female		
Variables	Mean	Std. Dev	N	Mean	Std. Dev	N	Mean	Std. Dev	N
Eight years of education or less[1]	89.50	16.35	383	90.94	16.31	158	88.49	16.35	225
High school education completed	104.74	18.77	406	104.41	18.38	169	100.98	19.08	237
Technical Formation Center or Professional Institute	108.84	19.60	43	109.61	21.31	20	108.16	18.45	23
Undergrad	116.47	15.86	73	117.63	15.17	28	115.75	16.40	45
Grad	128.79	12.11	16	124.99	17.21	4	130.06	10.59	12

Source: Study of reading habits and comprehension 2011

[1] Including those who did not complete high school

6 Reading Scores by log hourly wages

Illustration 2 Reading Scores by log hourly wages, Chile. (Reading behavior survey 2011)



7 Reliability coefficients

Reliability is the accuracy of measurement and allows us to quantify whether the set of items that make up the test actually measure what they say they do. From the logical point of view, reliability is defined as "the proportion of the variance of the results obtained in a measurement that is true variance." Reliability scale ranges from 0 to 1; therefore, the closer to 1, the more accurate the measurement.

For those older than 15, the reliability coefficient is 0.84, which, according to international agreements, is satisfactory or very satisfactory.

8 IRT

To calculate the scores we used the methodology of Item Response Theory (IRT). This methodology is composed of a variety of logistic models, where the main idea of the models is to calculation of participant probability with a latent trait structure and to correctly to answer an item subject to its characteristics, such as difficulty, discrimination, and chance.

Thus this methodology assumes the existence of an unobservable variable called a latent trait, which is to be estimated for each participant evaluated depending on the answers provided. Estimation of latent trait is known as the estimated participant ability.

The great advantage of this technique depends on the enforcement of assumptions, where the main one is the "unidimensionality," where the statistical dependence among the items can be explained by one dominant latent trait. In other words, the "unidimensionality" is centered on the test measuring one single dominant dimension, for example, mathematical ability, that does not disturb this ability when external factors are included, such as increasing difficulty in the construction of items that aim to measure mathematical but not reading ability..

Another assumption is the *local independence*, which is the underlying assumption of latent variable models. Defined by the concept of a relationship between variables, it specifies that the correct answer to an item X must be only influenced by the ability of the participant and the characteristics of the item and not by clues that may be delivered by the item X-1.

Finally, the last assumption of Item Response Theory is the "invariance" and is based on the stability of the estimated item parameters, i.e. independent of the distribution of latent traits where you get the same item parameters estimate.

Once generated, the final solution for those items with appropriate metric fit proceeded to estimate abilities of each of the participants evaluated in the test. Since the IRT has no direct interpretation, they are rescaled to facilitate interpretation. The scale used is centered at 100 and a standard deviation of 20.

9 Results

Table 5 Basic model and model including standardized test scores of reading comprehension

VARIABLES	(1) log of hourly earnings	(2) Select (mfx)	(3)[3] athrho	(4)[4] lnsigma	(5) log of hourly earnings	(6) Select (mfx)	(7) athrho	(8) lnsigma
Age	0.00583	**0.231******			0.00573	**0.228******		
	(0.00415)	(0.0241)			(0.00413)	(0.0243)		
Square Age		**-0.00260*****				**-0.00256*****		
		(0.000304)				(0.000306)		
Gender	**0.318*****	**0.874*****			**0.316*****	**0.870*****		
	(0.0968)	(0.115)			(0.0964)	(0.114)		
Marital Status		-0.0690				-0.0661		
		(0.132)				(0.131)		
Number of babies at home (< 2 years)		**0.405*****				**0.400*****		
		(0.138)				(0.138)		
Number of children at home (3-15 years)		**-0.170****				**-0.171*****		
		(0.0667)				(0.0664)		
Years of schooling	**0.104*****	**0.0540*****			**0.104*****	**0.0613*****		
	(0.0133)	(0.0177)			(0.0135)	(0.0195)		
Metropolitan	**0.118***	0.182			**0.118***	0.180		
	(0.0663)	(0.112)			(0.0661)	(0.112)		
years of work experience in the current job	**0.0319*****				**0.0319*****			
	(0.0114)				(0.0114)			
Square years of work experience in the current job	-0.000519				-0.000515			

[3] Athrho corresponds to the inverse hyperbolic tangent of rho, where rho is the covariance of the error terms between the selection and wage equation

[4] lnsigma corresponds to the log of sigma, where sigma is the variance of the residual.

	(0.000338)				(0.000338)			
Standardized score	**0.0671***				**0.0645***	**-0.0477**		
	(0.0368)				(0.0377)	**(0.0658)**		
Constant	**5.485*****	-5.328***	0.189	-0.444***	**5.491*****	**-5.334*****	0.181	-0.445***
	(0.347)	(0.431)	(0.242)	(0.0552)	(0.349)	(0.430)	(0.239)	(0.0548)
Observations	918	918	918	918	918	918	918	918

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 Basic model

VARIABLES	(9) log of hourly earnings	(10) Select (mfx)	(11) athrho	(12) lnsigma	(13) log of hourly earnings	(14) Select (mfx)	(15) athrho	(16) lnsigma
Age	**0.00631***	**0.120*****			**0.00656***	**0.138*****		
	(0.00372)	(0.0365)			(0.00378)	(0.0312)		
Square Age		**-0.00138*****				**-0.00161*****		
		(0.000424)				(0.000377)		
Gender	**0.212****	**0.987*****			**0.259****	**0.984*****		
	(0.102)	(0.130)			(0.125)	(0.133)		
Marital Status		-0.131				-0.169		
		(0.143)				(0.145)		
Number of babies at home (under 2 years)		**0.278***				0.169		
		(0.154)				(0.150)		
Number of children at home (3-15 years)		-0.113				**-0.154****		
		(0.0763)				(0.0754)		
Years of schooling	**0.102*****	**0.0488*****			**0.104*****	**0.0550*****		
	(0.0126)	(0.0176)			(0.0142)	(0.0182)		
Metropolitan	**0.129***	**0.208***			**0.132***	0.114		
	(0.0665)	(0.126)			(0.0678)	(0.125)		
years of work experience in the current job	**0.0301*****				**0.0252****			
	(0.0115)				(0.0115)			
Square years of work experience in the current job	-0.000465				-0.000356			
	(0.000336)				(0.000338)			
Standardized score	**0.0634***				**0.0700***			
	(0.0374)				(0.0384)			
Constant	5.618***	-2.913***	-0.0408	-0.506***	5.544***	-3.169***	0.0542	-0.490***

	(0.294)	(0.759)	(0.259)	(0.0481)	(0.369)	(0.592)	(0.364)	(0.0503)
Observations	730	730	730	730	736	736	736	736

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 Basic model

VARIABLES	(1) log of hourly earnings	(2) Select (mfx)	(3) athrho	(4) lnsigma	(5) log of hourly earnings	(6) Select (mfx)	(7) athrho	(8) lnsigma	(9) log of hourly earnings	(10) Select (mfx)	(11) athrho	(12) lnsigma
Age	0.00676*	0.145***			0.00667*	0.142***			0.00794**	0.130***		
	(0.00400)	(0.0316)			(0.00386)	(0.0310)			(0.00359)	(0.0425)		
Square Age		-0.00169***				-0.00165***				-0.00149***		
		(0.000386)				(0.000376)				(0.000538)		
Gender	0.283	0.979***			0.258**	0.969***			0.422**	0.920***		
	(0.206)	(0.138)			(0.127)	(0.133)			(0.164)	(0.155)		
Marital Status		-0.205				-0.185				-0.249**		
		(0.158)				(0.143)				(0.118)		
Number of babies at home (under 2 years)		0.173				0.168				0.211		
		(0.165)				(0.152)				(0.129)		
Number of children at home (3-15 years)		-0.156**				-0.158**				-0.138**		
		(0.0755)				(0.0765)				(0.0679)		
Years of schooling	0.108***	0.0388*			0.105***	0.0461**			0.117***	0.0364		
	(0.0164)	(0.0205)			(0.0140)	(0.0190)			(0.0142)	(0.0233)		
Metropolitan	0.136**	0.0888			0.133**	0.0948			0.148**	0.0706		
	(0.0675)	(0.125)			(0.0674)	(0.127)			(0.0710)	(0.121)		
indicator autonomy	-0.0173	0.0658							-0.00783	0.0534		
	(0.0223)	(0.0405)							(0.0228)	(0.0403)		
self-control indicator	-0.0105	0.0336							-0.0104	0.0286		
	(0.0346)	(0.0524)							(0.0355)	(0.0501)		
Locus control indicator	0.00171	0.0211							0.00369	0.0194		
	(0.0253)	(0.0462)							(0.0271)	(0.0463)		
years of work	0.0248**				0.0253**				0.0257**			

experience in the current job	(0.0115)				(0.0115)				(0.0113)			
Square years of work experience in the current job	-0.000334				-0.000363				-0.000400			
	(0.000346)				(0.000341)				(0.000346)			
Standardized score	**0.0722***				**0.0709***				0.0658			
	(0.0391)				(0.0403)				(0.0401)			
math skills indicator 1					-0.0199	0.0173			-0.0133	-0.0482		
					(0.0209)	(0.0442)			(0.0310)	(0.0565)		
math skills indicator 2					0.0167	0.0635			0.0318	-0.0343		
					(0.0280)	(0.0486)			(0.0458)	(0.0771)		
math skills indicator 3									-0.00341	0.0547		
									(0.0251)	(0.0398)		
Constant	5.574***	-3.644***	0.147	-0.488***	5.521***	-3.314***	0.0726	-0.491***	5.104***	-3.657***	0.736	-0.395***
	(0.691)	(0.633)	(0.702)	(0.0666)	(0.415)	(0.600)	(0.394)	(0.0508)	(0.562)	(0.770)	(0.650)	(0.143)
Observations	736	736	736	736	736	736	736	736	736	736	736	736

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1