I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13301

A Human Capital Index for the Italian Provinces

Alessandra Pasquini
Furio Camillo Rosati

MAY 2020

DISCUSSION PAPER SERIES

IZA DP No. 13301

A Human Capital Index for the Italian Provinces

Alessandra Pasquini
University of Rome "Tor Vergata"

Furio Camillo Rosati
CEIS and IZA

MAY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

A Human Capital Index for the Italian Provinces

Good health conditions and high quality education are crucial for children development and for their future contribution to the society. Human capital has been recognized as one of the crucial engines of economic growth. Nonetheless, it is often hard to establish a metric that allows to monitor its evolution and contribute to assess the effects of policies. In Italy, the use of such an index at national level may not be enough to have a clear picture of the human capital conditions. Socio-economic characteristics and public services are highly heterogeneous across the Country. There is, therefore, good ground to believe that also the human capital presents substantial differences across the Italian Provinces. To take such a high heterogeneity into consideration, we develop a Human Capital Index for Italy disaggregated at provincial level. The results show very large differences across Italian Provinces in terms of human capital, mostly driven by the variation in the quality of educational. Strikingly, the differences among Italian Provinces span a range that goes from best performers among high income countries to middle and low income countries. Finally, we classify the Italian Provinces in three main clusters according to their HCI and show how the clusters differ in terms of several socio-economic characteristics.

JEL Classification: I20, I24, I28

Keywords: human capital, index, Italian provinces

Corresponding author:
Furio Camillo Rosati
Department of Economics and Finance
University of Rome Tor Vergata
via Columbia n. 2
00133 Roma
Italy

E-mail: f.rosati@economia.uniroma2.it

1. Introduction

Italy is far from being a homogeneous country and it is characterized by large differences in the economic and social structure. The “Questione Meridionale”[1] has been at the center of the political debate since the reunification of Italy in 1861. The last Italian National Institute of Statistics (ISTAT) report (Istat, 2019) on the differences in terms of well-being across the Italian territory illustrates it very clearly in terms both of labour market and of economic conditions. In 2016, the employment rate in the northern area of the country was 50% higher than that of the South (i.e., 70.6% vs 47%). In 2012, the disposable income by family was about 55.000 euros in the Northern Province of Milan and about 25000 euros in the southern Province of Enna. The report also evidences differences in the local administration capabilities. In 2015, the revenues of the municipalities covered on average 19% of their expenses in the North and 6.1% in the South. This difference was even higher at provincial level. In the Province of Milan, the municipalities were able to cover 35.6% of their expenses, while in the Province of Catania only 4.7%. Beyond the North-South division, there are other striking differences within the Italian Provinces. In 2015, the rate of deadly and permanent disability work related accidents ranged from 5.3 every 10’000 workers in the Sicilian Province of Messina to 53.2 in the Sicilian Province of Siracusa. With respect to security, the number of reported crimes ranged from 484.6 every 10’000 residents in the northern Province of Rimini to 63.2 in the southern Province of Potenza. The percentage of prisoners with respect to total jail capability in 2016 ranged from 24.8% in the central Province of Arezzo to 180.1% in the northern Province of Como. The Provinces present large differences with respect to public services as well: for example, the offer of seats per kilometre in public transport ranged from 16’218 in Milan to 293.8 in Ragusa in 2015.

The “New growth theories” see the human capital as one of the main engines of growth and as one of the reasons that can explain persistence in the differences in growth rates. Therefore, in an heterogenous country like Italy, it is of interest to assess the existence of large geographical differences in the stock of human capital, as it may contribute to understand the reasons behind the gap among different areas of the country and to monitor and identify appropriate intervention policies.

[1] The expression "southern question" indicates the set of problems posed by the existence in the South of Italy from 1861 until today of a lower level of economic development, of a different and more backward system of social relations, of a more weak development of many important aspects of civilian life compared to the central and northern regions.

To this aim, we estimate the Human Capital Index proposed by the World Bank[2] at the provincial level over the Italian territory. This index brings together both the education and the health components that are typically assumed to constitute human capital. It uses a single metric for a large number of Countries requiring the adoption of a set of simplifying assumptions and cross-country standardizations. Nonetheless, it allows us to compare the provincial rankings with the world ranking provided by the World Bank. The comparison gives us an idea of the magnitude of the gap across provinces.

Our results evidence a high heterogeneity in the human capital across the country, mostly driven by the differences in educational quality. The comparison proofs even more the heterogeneity characterising Italy. Indeed, the HCI of the 110 Italian Provinces spans over 80 positions of the world ranking. The human capital index is correlated with other socio-economic variables, such as income and access to some public services and we could identify three clusters within the Italian Provinces, each with different characteristics in terms of labour market, childhood educational and social services and health conditions. We also show the robustness of the ranking across Provinces with respect to several different assumptions relative to the measurement of its components.

The rest of the paper is organised as follows: in section 2, we describe the World Bank Human Capital Index and the data we use to build it at provincial level, and we provide some descriptive statistics. In section 3, we show the distribution of the Human Capital Index across Provinces, compare the Italian ranking with the world's ranking and investigate the index relationship with other socio-economic characteristics. In section 4, we provide some robustness checks and section 5 we draw some final conclusions and underline the path that can be followed in the future.

2. Methodology and Data

In order to insure the comparability with the national and international data available, we build the human capital index following the methodology proposed by the World Bank (Kraay, 2018). Its components are education and health. The methodology has several limitations, because the index must be replicable worldwide. For example, the human capital index does not include information on the tertiary education, latent health is measured with imperfect proxies and the educational and health returns are world averages (which may not be representative for some countries if these returns are highly heterogeneous). At country level, some of the limitations could be overcame

[2] https://www.worldbank.org/en/publication/human-capital

using a country-tailored human capital index. Nonetheless, the use of a more refined index, exploiting the availability of more detailed statistics, would not allow us to compare the results of the Italian Provinces with those of the rest of the world. We would have a measure of the differences in the index among Provinces, without the possibility of performing any international comparison.

2.1. The World Bank Human Capital Index

The World Bank's human capital index is built following the latent variable approach (Folloni and Vittadini 2010). It aims to offer a metric of the human capital stock with which a young person enter adulthood and, possibly, the labour market. Therefore, education and health are its two components. They are aggregated using the respective returns in the labour market and monetized using an exponential function. The potential human capital stock measured in this way is then weighted by the probability of survival of a child to school age (age 5). Finally, the weighted monetarized value is normalized with respect to a benchmark reflecting the best possible performance. Thanks to this adjustment, the index ranges between 0 and 1.

The expected human capital of a child (using the same notation as in Kraay, 2018) is then given by:

$$h_{NG} = pe^{\Phi s_{NG} + \gamma z_{NG}}$$

where p measures the probability that a child survives to age 5, s_{NG} is a measure of the average education level achieved and z_{NG} the measure of average health conditions. Φ and γ are, respectively, education and health returns. The value of these two parameters was selected in Kraay (2018) on the base of the existing evidence. This author set the coefficients, respectively, to 0.08 and 0.65. We follow his choices, but we will perform some robustness checks.

The final form of the human capital index, normalized with respect to the benchmarks, is given by:

$$HCI = \frac{p}{p^*} \times e^{\Phi(s_{NG} - s^*)} \times e^{\gamma(z_{NG} - z^*)}.$$

The value p^* measures the best possible outcome in terms of children survival (i.e., it is equal to 1). s^*measures the best possible performance in terms of education (i.e., the level of education that would be obtained if all children would complete higher secondary school and benefit from the highest observed quality education). Finally, z^* measures the best possible performance in terms of health (i.e., it corresponds to a situation where all children are expected to have the highest possible level of health as adults).

2.2 Data and variable definition

We measure the child probability of survival to age 5 with the survival rate of children between 0 and 4 years old. The data on children survival rate by Province were obtained from the mortality tables of the Italian National Institute of Statistics (ISTAT).

Table 1 presents the survival rate of children aged 0-4 by geographical macro-area. As it is evident, the survival rate does not present significant differences across macro areas.

Table 1 - Survival Rate of children aged 0-4 by geographical macroarea. Reference year: 2016. Source: ISTAT.

MACROAREA	SURVIVAL RATE
Centre	0.999
Islands	0.998
North East	0.999
North West	0.999
South	0.999

As suggested in Kraay (2018), we measure the average education level achieved s_{NG} as the expected learning-adjusted years of school (Filmer et al. 2018).

Children might attend school for the same number of years but accumulate different level of human capital because of the different quality of the school they attended. Learning-adjusted years of school (LAYS) combines information about both years of schooling and learning achievements. It is defined as:

$$Expected\ LAYS_{i,t} = Y_{i,t} \frac{L_{i,t}}{\max_i(L_{i,t})}$$

where $Y_{i,t}$ is a measure of the average years of schooling computed as the sum of net school enrolment rate for each grade in Province i at time t. $L_{i,t}$, is a measure of the level of learning achieved by children, defined as the average test score obtained by children. School achievements are normalized with respect to the highest test score observed in the sample.

To build the expected LAYS we need data on both the net enrolment rate by grade ($Y_{i,t}$) and children test scores ($L_{i,t}$).

The net enrolment rate by grade is defined as total number of students in the expected age group for a given level of education enrolled in that level, expressed as a percentage of the total population in that age group. The data on the number of children enrolled in school are from the Ministry of Education. Unfortunately, this data do not include information on the autonomous Provinces of Trento and Bolzano and on the autonomous region of Valle d'Aosta. Moreover, they do not provide information on a new type of vocational high schools, called IeFP, which are managed from the provincial administrations. Note that for these schools also the National Institute for the Evaluation of the Education and Training System (INVALSI) tests are not available. Therefore, we had to exclude these schools from our calculations. The exclusion of the group of children attending vocational schools from the HCI is equivalent to assuming that their human capital is identical to that of the children who dropped out after lower secondary school. This will generate a downward bias in the HCI, which will be larger the larger is the human capital accumulated through vocational schools. This may distort the relative ranking of Provinces if the participation to vocational schools differ substantially across them and it is quantitatively relevant. For the child population, we used ISTAT data on the population resident by Province on the 1st January of each year t. This data was associated to the academic year spanning from t-1 to t (i.e., for the academic year 2014/15 the residing population referred to 1st January 2015).

Figure 1, reports the expected average years of schooling built using the net enrolment rate by Province.

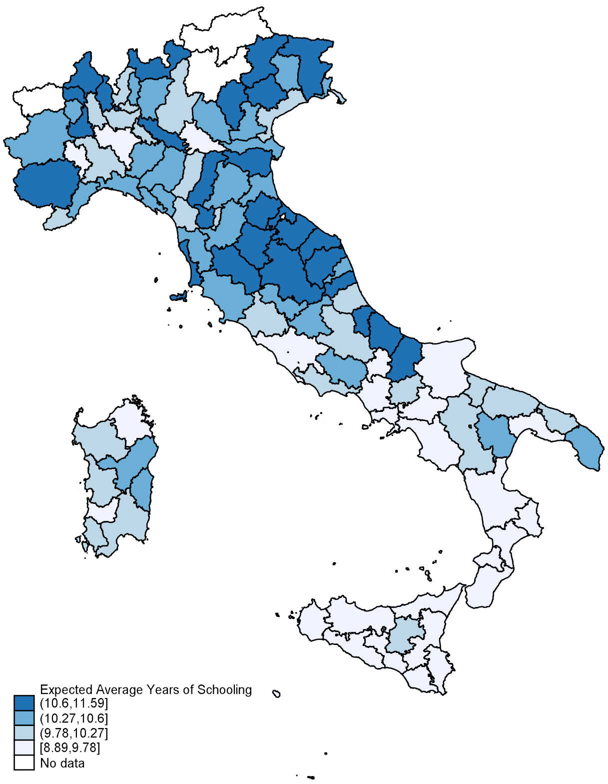

Figure 1 - Expected average years of schooling by Province. To build this measure the net enrolment rate was used. Reference year: 2016. Source: authors' elaboration on data from ISTAT and Ministry of Education.

The best and worst performing Provinces differ in terms of expected years of schooling by 2-3 years. The lowest values are concentrated in the South of Italy. The highest values, instead, are concentrated in centre-eastern and in the northern Provinces. If all children went to school for all years, the expected years of schooling would be equal to 13. Even the best performing Province is more than one year behind this theoretical maximum because of repetition and drop out.

Data on test scores were provided by the INVALSI. Each year, INVALSI run a nationwide learning test. The tests, relative both to mathematics and to reading, are submitted to children enrolled in the second and fifth grade of primary school, the third grade of lower secondary school and the second grade of upper secondary school. As suggested in Kraay (2018) we used the data on the highest grade available (i.e., the second year of upper secondary school, which corresponds to 10th grade) averaging between mathematics and reading test scores. We corrected the test scores to consider the possible cheating. The correction factor is calculated directly by INVALSI following Quintano et al. (2009). We normalized the test scores to range between 0 and 100. The best performing Province was Lecco. To build the final expected LAYS, we used this Province as a benchmark.

In figure 2, we report the average normalized mathematics and reading test scores by Province.

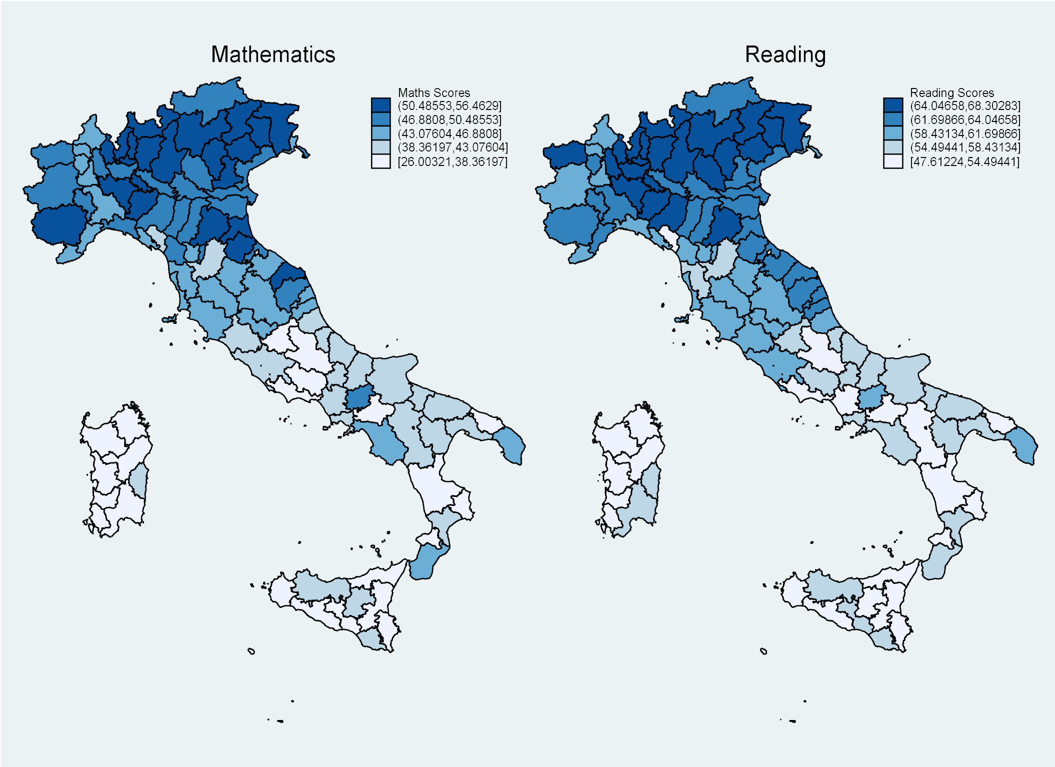

Figure 2 - Average Mathematics and Reading test scores by Province. Values are normalized to range between 0 and 100. Reference year: 2016. Source: authors' elaboration on INVALSI data.

As it is possible to see from the figure, the average test scores are highly heterogeneous across the country. Provinces in the South and in the Islands present the lowest average scores. Central Provinces have medium scores and Northern Provinces have the highest scores. The relative performances are very similar between mathematics and reading. Few Provinces make an exception. In figure 3, we report the learning-*adjusted years of schooling* built using the net enrolment rate *by* Province. Considering the test scores, further increases the distance between the theoretical maximum of 13 years and the value observed. The difference across provinces is substantial.

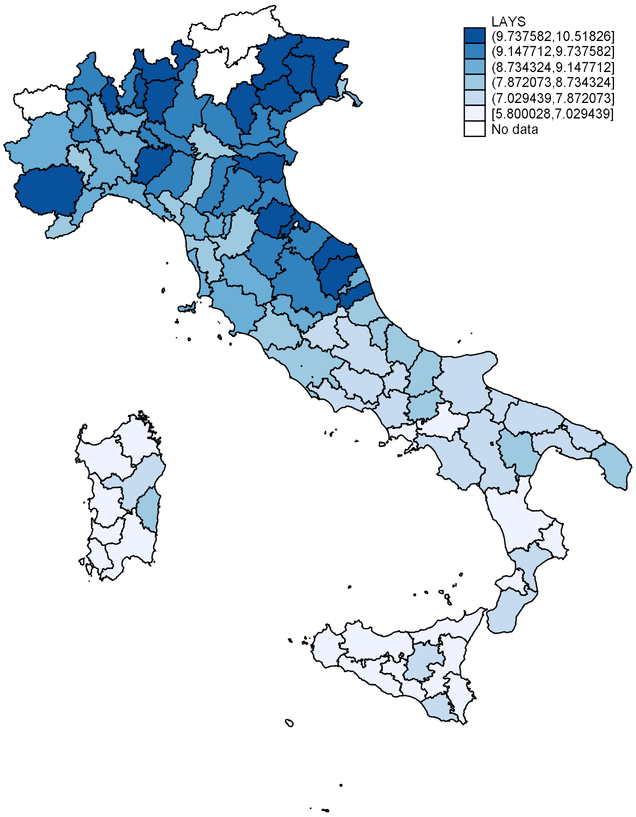

Figure 1 – Learning-adjusted years of schooling by Province. To build this measure the net enrolment rate was used. Reference year: 2016. Source: authors' elaboration on data from ISTAT and Ministry of Education.

The best and worst performing Provinces differ in terms of LAYS by 4-5 years. The lowest values of LAYS are concentrated in the South of Italy and in the Islands. The highest values, instead, are concentrated in centre-eastern and in the northern Provinces.

The average health conditions z_{NG}T should summarize all aspects of health conditions that have an influence on productivity. Weil (2007) and Kraay (2018) propose to use adult survival rates[3]. As for the coefficient measuring the returns of adult survival rates, we have used those proposed by Kraay (2018). Data on adults (i.e., individuals aged between 15 and 60 years old) survival rate by Province were obtained from the mortality tables of ISTAT.

[3] Weil (2007) and Kraay (2018) also propose stunting rate as a proxy for health condition. To the best of our knowledge, no data are available for Italy on this variable. We refer the reader to Kraay (2018) for a discussion on this approach.

In table 2, we report adult survival rate by geographical macro-area.

Table 2 - Adult survival rates by geographical macro-area. Reference year: 2016. Source: authors' elaboration on ISTAT data.

AREA	SURVIVAL RATE
Centre	0.949
Islands	0.940
North East	0.950
North West	0.948
South	0.944

3. Results and discussion

3.1. The Human Capital Index by Province

In figure 4, we present Human Capital Index in 2016 across the Italian Provinces (Appendix A contains the data for all the Provinces). The most recent available data refer to 2017. However, in 2017 ISTAT used the administrative provincial division of Sardinia established in 2016, while MIUR and INVALSI continue to use the previous division. Therefore, in 2017 we would have some missing provinces in that region. For this reason, we prefer to present the data relative to 2016. However, we show that the ranking of provinces between 2016 and 2017 does not changes in any substantial way for the rest of Italy.

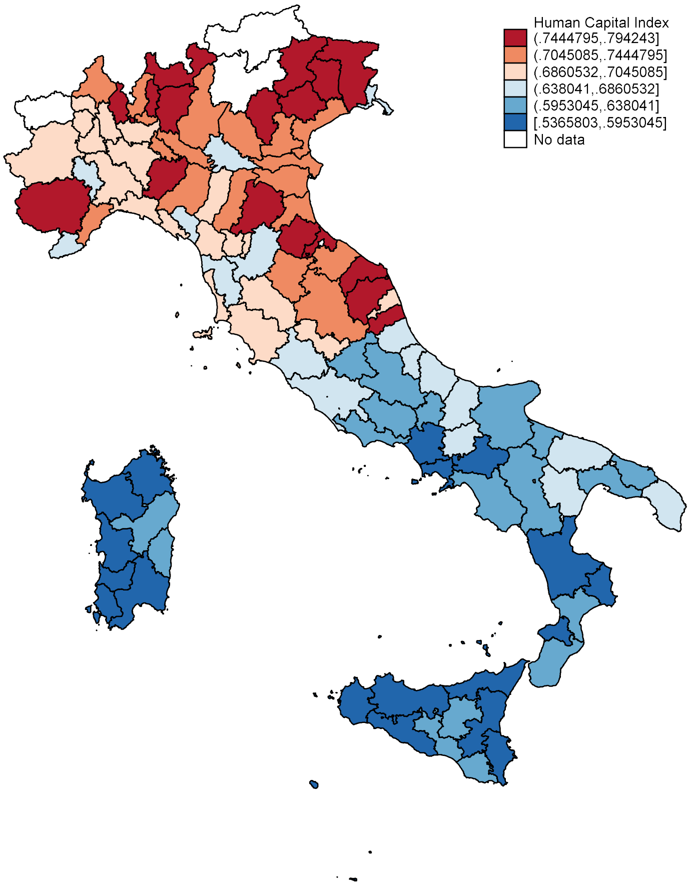

Figure 4 - Human capital index by Province. Reference year: 2016.

The Italian Provinces are concentrated in the upper part of the index range: all of them presents a value of the human capital index higher than 0.5 (the index ranges between 0 and 1). HCI values are highly heterogeneous across the country. This testify how relevant it is the calculation of an index disaggregated at provincial level. The highest values of the index are observed for the Northern Provinces, in particular in the North-East. The worst performers are found in the Islands and in the South. The dispersion is large, e.g., the difference between the best and the worst performer is about 48 per cent.

In table 3, we present the HCI of the five best and worst performers (see appendix A for the complete ranking): most of the best performers are in Emilia-Romagna or Lombardy, while most of the worst performers are in Sardinia.

Table 3 - Best and worst performing Provinces according to their HCI. Reference year: 2016.

RANK	PROVINCE	REGION	MACRO-AREA	HCI
1	LECCO	LOMBARDIA	NORTH WEST	0.79
2	FORLÌ-CESENA	EMILIA ROMAGNA	NORTH EAST	0.79
3	SONDRIO	LOMBARDIA	NORTH WEST	0.79
4	PIACENZA	EMILIA ROMAGNA	NORTH EAST	0.78
5	UDINE	FRIULI-VENEZIA G.	NORTH EAST	0.77
104	MESSINA	SICILIA	ISLANDS	0.56
105	CROTONE	CALABRIA	SOUTH	0.56
106	OLBIA-TEMPIO	SARDEGNA	ISLANDS	0.56
107	CARBONIA-IGLESIAS	SARDEGNA	ISLANDS	0.55
108	MEDIO CAMPIDANO	SARDEGNA	ISLANDS	0.54

In order to assess how the different components affect the HCI, we present, in figure 5, the correlation between the human capital index and its components for 2016.

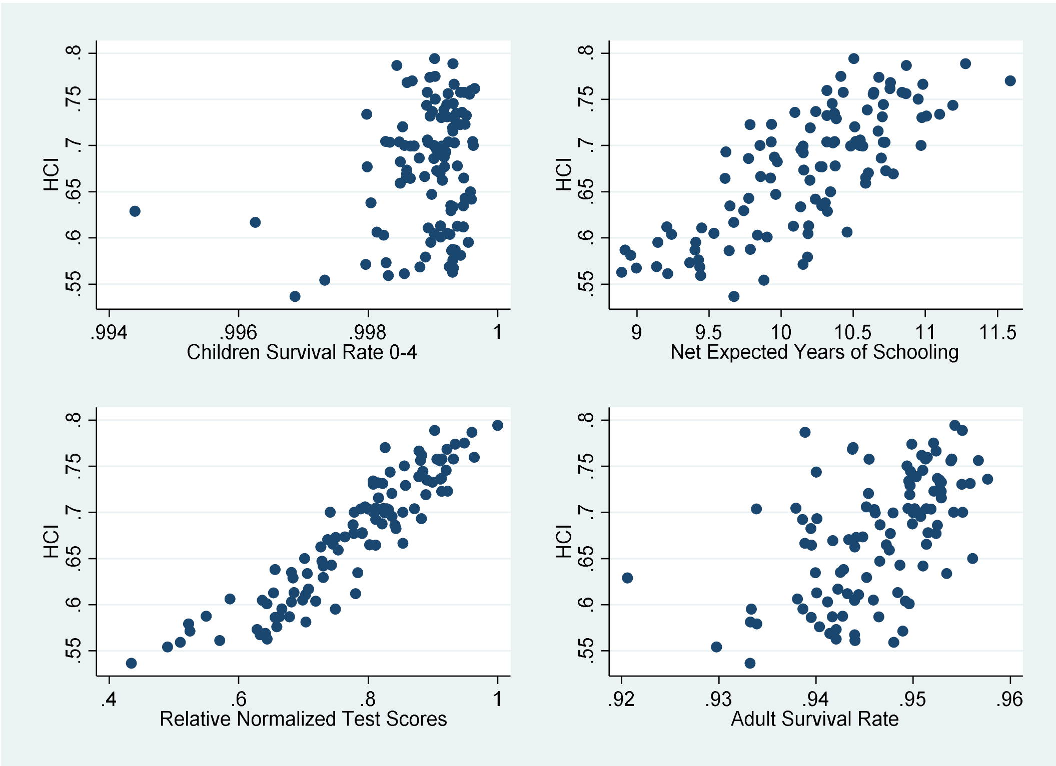

Figure 2 - Correlation between the HCI and its components. Reference year: 2016.

The HCI has an extremely low correlation with the children survival rate. This is not surprising given that the values of the latter variable are very similar across Provinces. The HCI presents the highest correlation with the ratio of the normalized test scores, followed by the expected years of schooling and the adult survival rate. This is confirmed by table 4, where we display the correlation between the Human Capital Index and its components.

Table 4 - Correlation between the HCI and its components. Reference year: 2016.

Component	Correlation with HCI
Children Survival Rate	0.215
Net Expected Years of Schooling	0.778
Normalized Test Scores	0.923
Adult Survival Rate	0.569

3.2. Recent dynamics

In figure 6, we look separately at the absolute change over time of the HCI by Provinces above or below the national average.

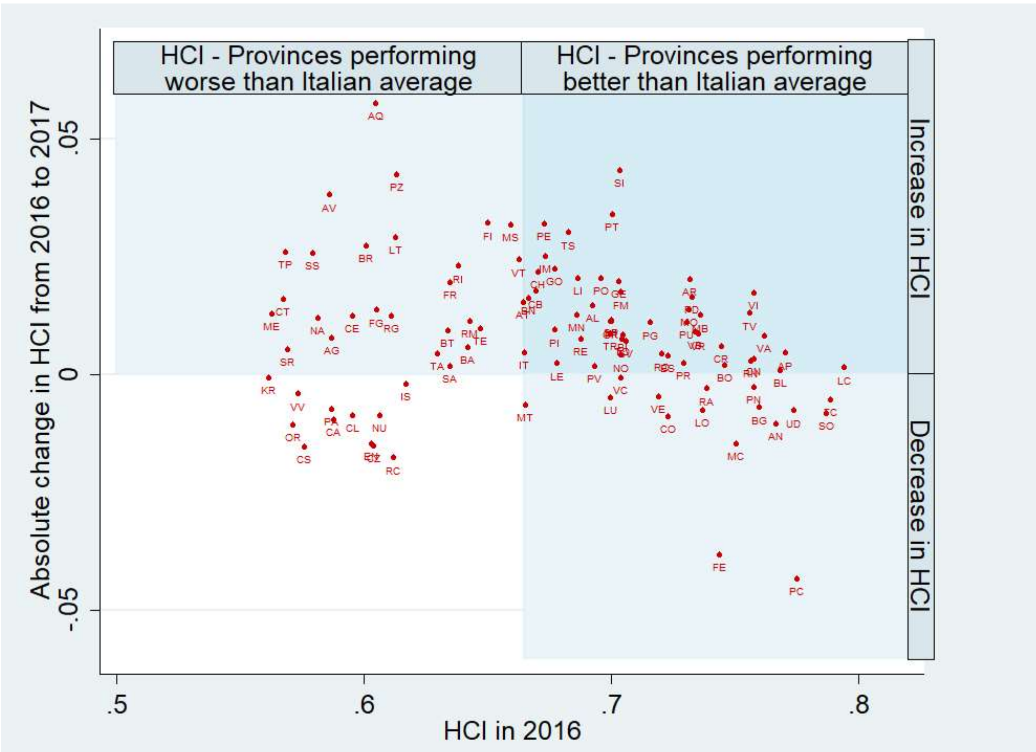

Figure 6 - Changes in the HCI between 2017 and 2016 by province average performance.

Most of Provinces’ HCI in 2017 improved with respect to 2016. For the minority of Provinces that saw a deterioration, the reduction in the HCI was rather small: the two exceptions are the Provinces of Ferrara and Forlì-Cesena, which experienced the largest reduction (interestingly, both Provinces belong to the Emilia-Romagna region). The Provinces with the larger improvement are mainly Provinces characterized by a human capital index lower than the Italian average in 2016. In fact, most of the Provinces having a human capital index higher than the Italian average had an improvement close to zero. Nonetheless, all the variations between 2016 and 2017 are small, i.e., the province of L’Aquila experienced the highest variation, which amounted to 8% of the Italian HCI average in 2016. This suggests that the worst performing Provinces are converging, but only slowly, to the best performing ones[4].

3.3. The Human Capital Index of Italian Provinces in a global perspective

In order to put the HCI, and especially its dispersion at provincial level, in perspective, we compare the HCI index of the Italian Provinces with that of the different countries in the world computed by the World Bank.

To this aim, we had to harmonize our measure with that utilized by the World Bank. The World Bank built the Human Capital Index using, as the preferred choice, the TIMSS test for mathematics and the PISA test scores for reading. To compare the HCI of the Italian Provinces with that of other countries, we should use the same test scores. Unfortunately, TIMSS and PISA data are not available for Italy at Province level. To address this issue, we use, following Kraay (2018), a conversion factor. Additional information on the harmonization are available in appendix C and in Kraay (2018).

In figure 7, we present the results of the comparison. Countries are plotted in blue while Italian Provinces are plotted in red.

[4] Note that, due to data issues, in 2017 the HCI for the Provinces of Sardinia (some of the worst performing Provinces in 2016) is not available.

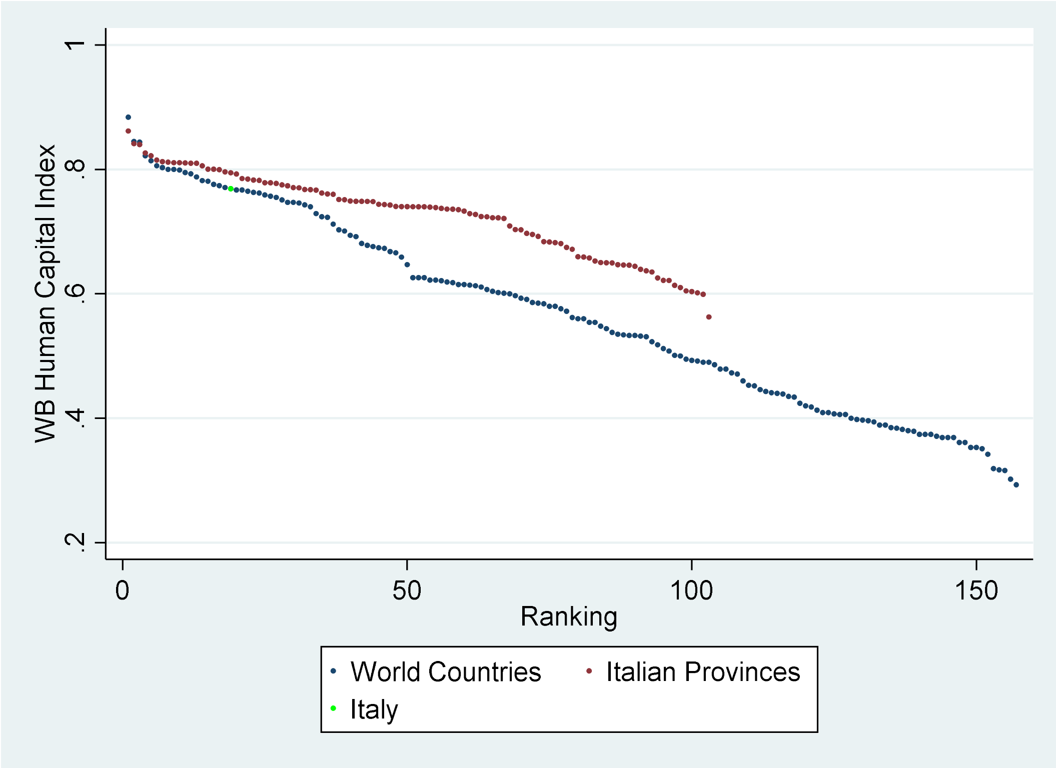

Figure 7 - Comparison of Italian Provinces and world countries HCI values. Data on the world countries HCI were taken from World Bank. Reference year: 2016.

The variation of the HCI across the Italian Provinces span a large portion of worldwide variation. While the HCI of the best performing Italian Provinces is equal to that of the best performing countries, the worst performing Italian Provinces set themselves below some middle-income countries such as Thailand, Mexico and Iran. More in detail, the HCI of Lecco, the Italian best performing Province, set itself between the two best performing world countries: South Korea and Singapore. The HCI of Vibo Valentia, the Italian worst performing Province according to the harmonized scores, is approximately equal to that of Jordan, ranked 79th among world countries.

3.4. Income per capita and Human Capital

A higher level of income per capita is likely to be associated with a higher level of HCI. At a macro-economic level, higher levels of income per capita are likely to be associated with higher levels of local fiscal revenues. Higher fiscal revenues are likely to be associated with a higher level of the services provided in terms of education and health. At a micro-economic level, better economic conditions of the parents are often associated with better educational performances of children and better health conditions[5]. In figure 8, we plot the relationship between the HCI and the income per capita observed for the Italian Provinces. Income per capita is measured as pre-tax income and data are from the taxable income files provided by the Ministry of Economy and Finance.

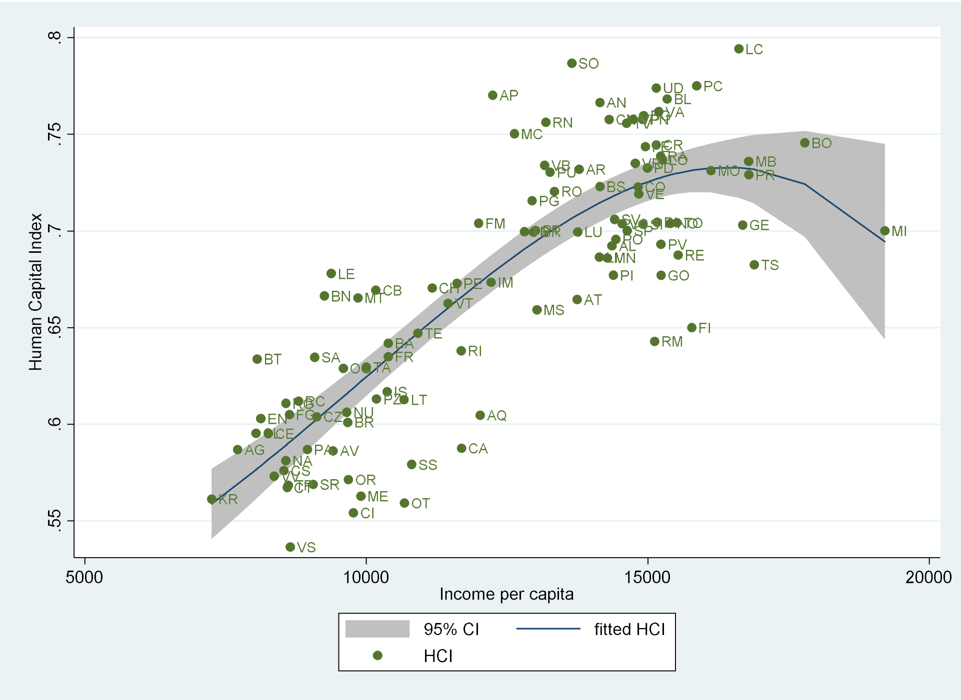

Figure 8 - Human Capital Index with respect to income per capita. Reference year: 2016.

Low income per capita is associated with low values of the HCI. Above a certain level of income, however, as income grows, the increase of the HCI abate. In particular, the Provinces seem to cluster into two different groups. The first group is characterized by both low values of income per capita (ranging between 7500 and 12500 Euros) and low values of the human capital index (ranging between 0.5 and 0.7). The second group is characterized by higher values of per capita income and values of the human capital index which are higher but proportionally lower. An interesting result arising from the figure is also that a large number of the most populated Provinces (such as Rome, Milan, Turin, Florence and Genoa) present values of the human capital index lower than expected according to their income (as we can see from the fact that they are located below the fitted blue line).

[5] See, among others, Corak, 2013 and Blane, 1995

3.5. The Human Capital Index and the Provinces' characteristics

In this section, we try to identify groups of Provinces that present similar HCI and to look at some of the characteristics associated with the different groups. We first use cluster analysis to divide the Provinces into different clusters according to their human capital index. Then, we show how the groups differ along some characteristics that are likely to be correlated with the human capital index.

3.5.1. Cluster Analysis

To identify the different groups of similar Provinces we have used cluster analysis. We employ Euclidean distance as a dissimilarity measure and the k-means algorithm. We choose the number of clusters to use according to the distribution of the Human Capital Index displayed in appendix BFigure . As the distribution appears to be tri-modal, we use three clusters.

In table 5, we present the average of the Human Capital Index for the three clusters. It is possible to see that the average values coincide fairly well with the three-modal points of the distribution (figure B.1.). The number of Provinces in each group is similar.

Table 5 - Clusters' characteristics. Clusters number was established according to the distribution of Provinces' Human Capital Index. Clusters were chosen using a k-means algorithm and the Euclidean distance as a dissimilarity measure. Reference year: 2016.

Cluster	Names	Number of Provinces	Average HCI
1	Low HCI	36	0.59
3	Medium HCI	39	0.75
2	High HCI	33	0.68

In figure 9, we show the clusters to which each Province belongs.

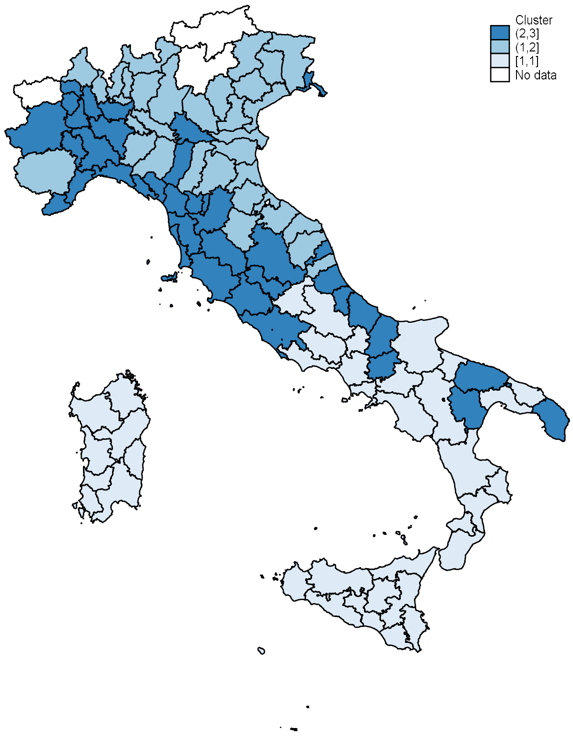

Figure 9 - Provinces by cluster of belonging. Clusters number was established according to the distribution of Provinces' Human Capital Index. Clusters were chosen using a k-means algorithm and the Euclidean distance as a dissimilarity measure. The first cluster

Provinces belonging to the low HCI cluster are concentrated in the South and in the Islands. Provinces in the medium HCI cluster are instead in the Centre and in the north-western areas (with some exceptions in the South). Finally, Provinces in the high HCI cluster are in the north-eastern area.

We compare the distributions of some socio-economic characteristics of the three groups. In figure 10, we show the kernel density functions of the 3 clusters for 4 different characteristics of the labour market. In the first graph from the upper left corner, we display the distribution of the employment rate, in the second, that of the inactivity rate and in the bottom half the distribution of the share of workers employed in the industry and building sector and that of the share of worker employed in the service sector.

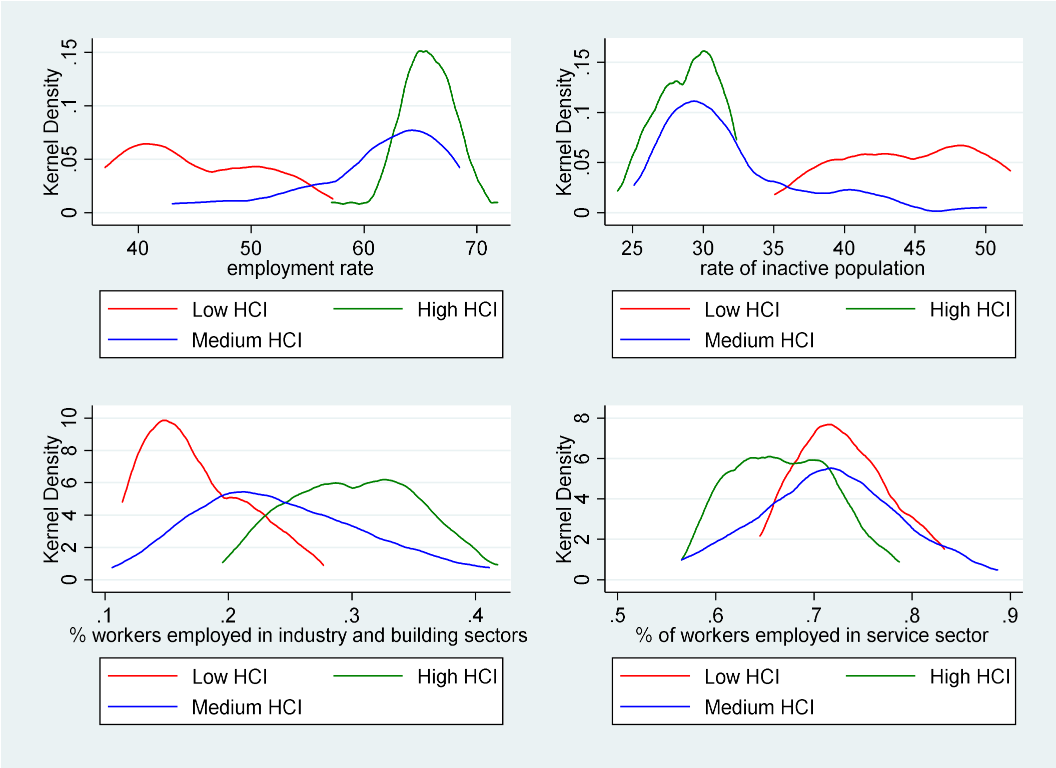

Figure 10 - Distributions of labour market characteristics by cluster. Reference year: 2016. Source for Provinces' characteristics: ISTAT.

The level and the structure of employment appear to be associated with the HCI. Provinces belonging to the high and medium HCI cluster are characterized by higher values of the employment rate and lower values of the rate of inactive population, while the opposite is true for Provinces belonging to the low HCI cluster. The occupation of the high HCI cluster is more concentrated in the industry and building sectors with respect to the medium and the low HCI clusters, and less concentrated in the service sector. The low and the medium HCI clusters do not differ much in terms of the share of workers employed in the service sector.

In figure 11, we display the kernel densities of the 3 clusters according to measures of the services available in the Provinces for early childhood education and families' social assistance. In the first graph we provide the density for the percentage of individuals benefitting of socio-educational services for early childhood with respect to the target population and in the second the average municipality expenditure for kindergarten per child aged 0-2. The Provinces belonging to the low HCI cluster appear to be characterized by a substantially lower levels of access to early childhood educational services and of kindergarten per capita expenditure than the others.

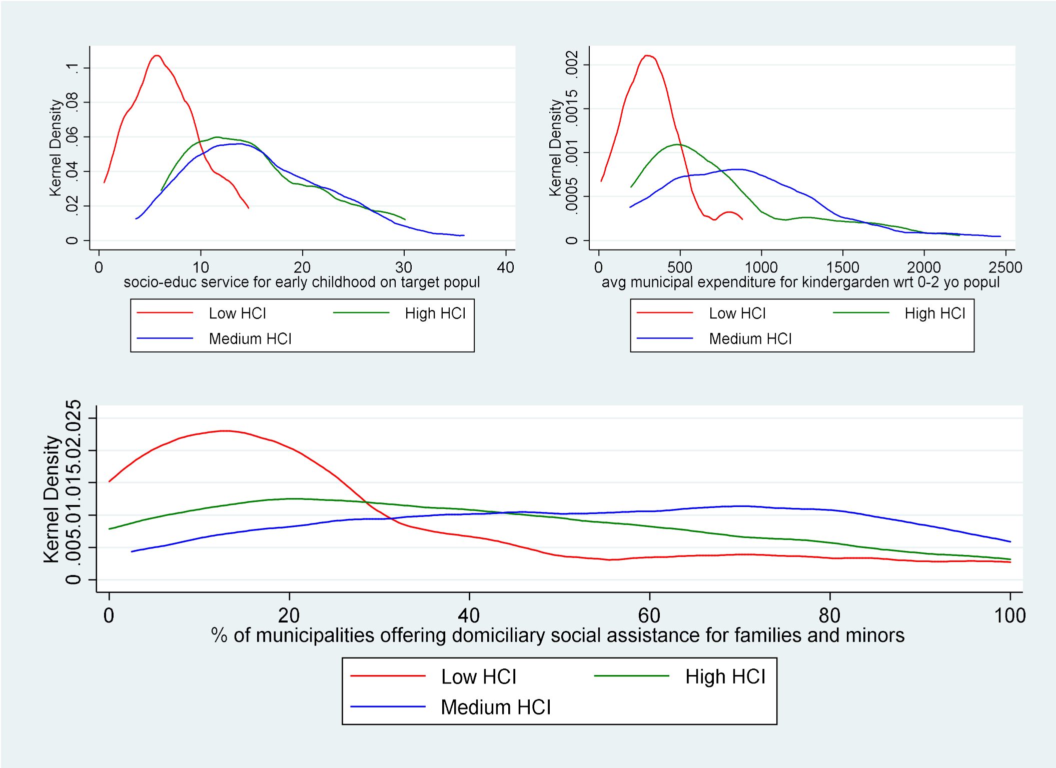

Figure 11 - Distributions of childhood education and social assistance characteristics by cluster. Reference year: 2016. Source for Provinces' characteristics: ISTAT.

Finally, in the bottom part of the graph we present the percentage of municipalities offering domiciliary social assistance to families and minors. Again, the Provinces in the low HCI cluster are characterized by lower percentages of municipalities providing assistance.

In figure 12, we display the kernel densities for 4 variables relative to some health characteristics: the share of deaths due to a different set of diseases. The 3 clusters differ in terms of health as well. In particular, the low HCI cluster presents lower values of the incidence of deaths due to infectious and parasitic diseases, nervous system and sense organs diseases and respiratory system diseases. Instead, they present a higher percentages of deaths due to endocrine, nutritional and metabolic diseases. The distributions are fairly similar between the medium and the high HCI clusters.

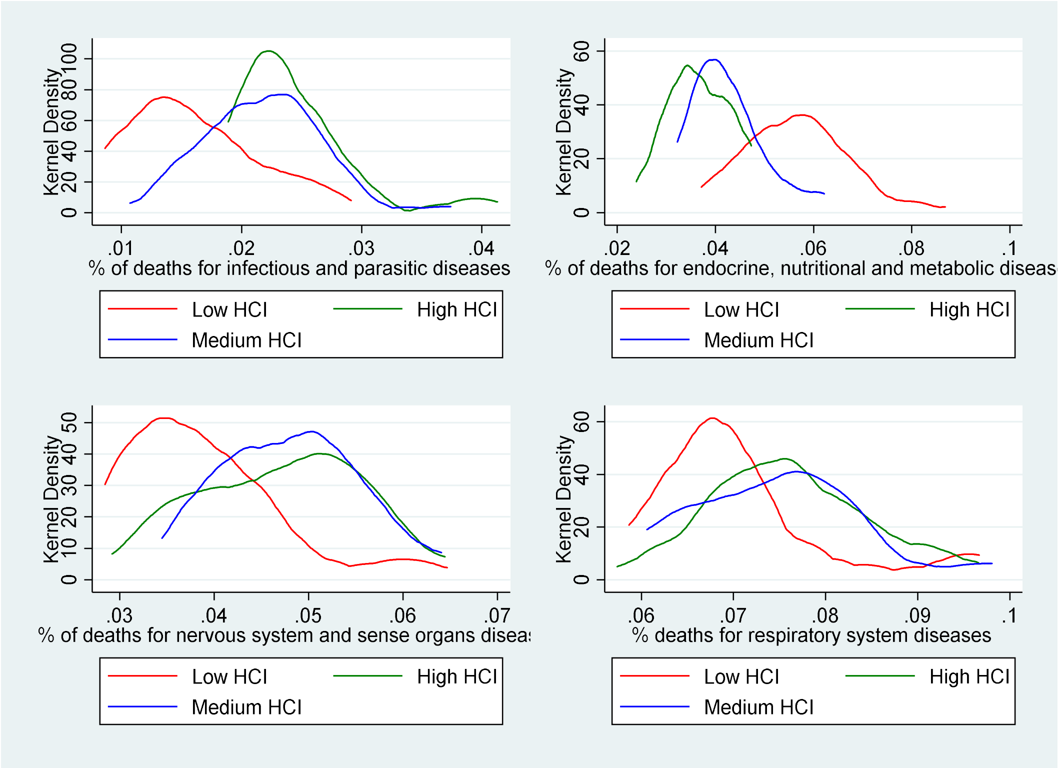

Figure 3 - Distributions of health characteristics by cluster. Reference year: 2016. Source for Provinces' characteristics: ISTAT.

To conclude, the high and medium HCI clusters have similar characteristics in terms of health and social assistance, nonetheless they are different in terms of labour market (especially with respect to the sectors of employment), and in terms of early childhood education. The low HCI cluster, instead differ from the other two cluster with respect to all characteristics.

4. Robustness Checks

Building the HCI required a series of assumptions. In this section, we check whether relaxing them affects the value and the rankings of the HCI. We look at the robustness of the results with respect to different values for the returns to health and education, to different methods of test scores aggregation and, finally, to the use of gross instead of net enrolment rates.

For the returns to education and to health conditions, we have chosen the same values as in Kraay (2018) who followed the international literature on these topics (see section 2.1.). In figure 13 we check how much the results are sensitive to the particular selection of returns considered. In Figure 19, we report the HCI by Province computed with two different set of returns. The left hand side present the HCI computed setting $\Phi = 1$ and $\gamma = 1$, the right hand side, instead, show the HCI computed with the parameters utilized for the estimates presented above (i.e., $\Phi = 0.08$ and $\gamma = 0.65$).

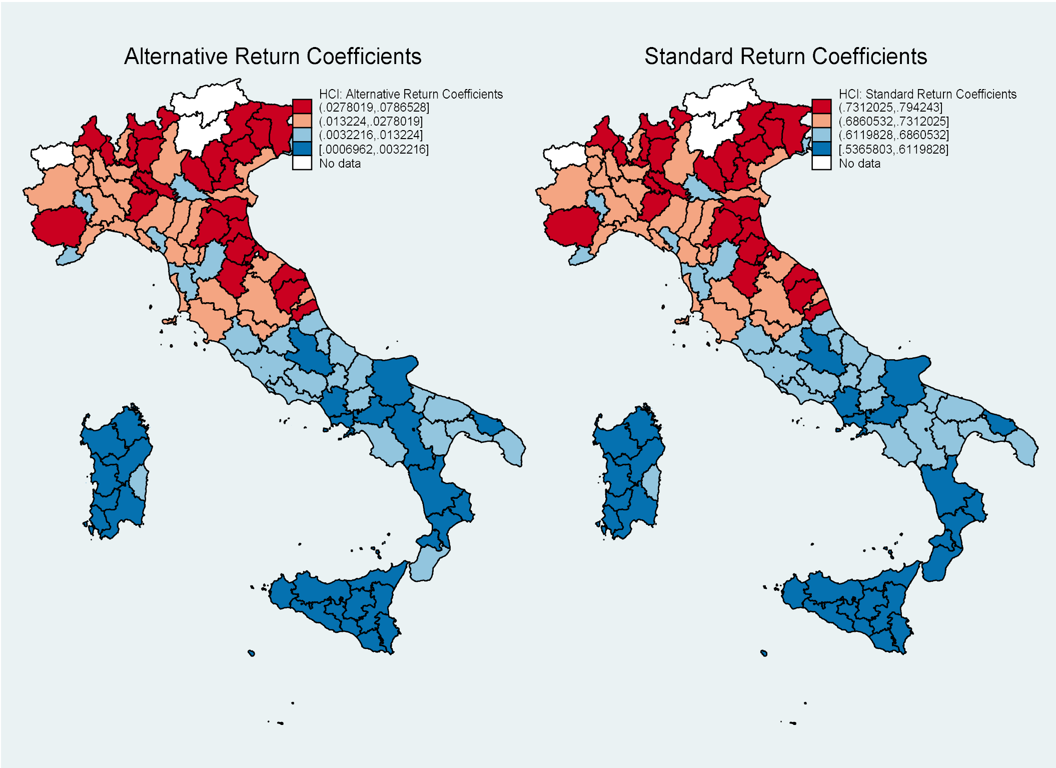

Figure 13 - Robustness check with respect to different returns of education and health. On the left hand side the HCI was calculated setting the return coefficients as Φ=1 and γ=1. On the right hand side they were set as Φ=0.08 and γ=0.65. Reference year: 2016.

While the absolute value of the HCI not surprisingly changes substantially, the relative ranking does not, indicating that the choice of the returns to health and educations is not likely to affect substantially the relative ranking of Provinces in terms of HCI.

Figure 14 offers additional evidence on the robustness of the rankings to different coefficients choices. The ranking obtained using the two sets of parameters discussed above are plotted against each other.

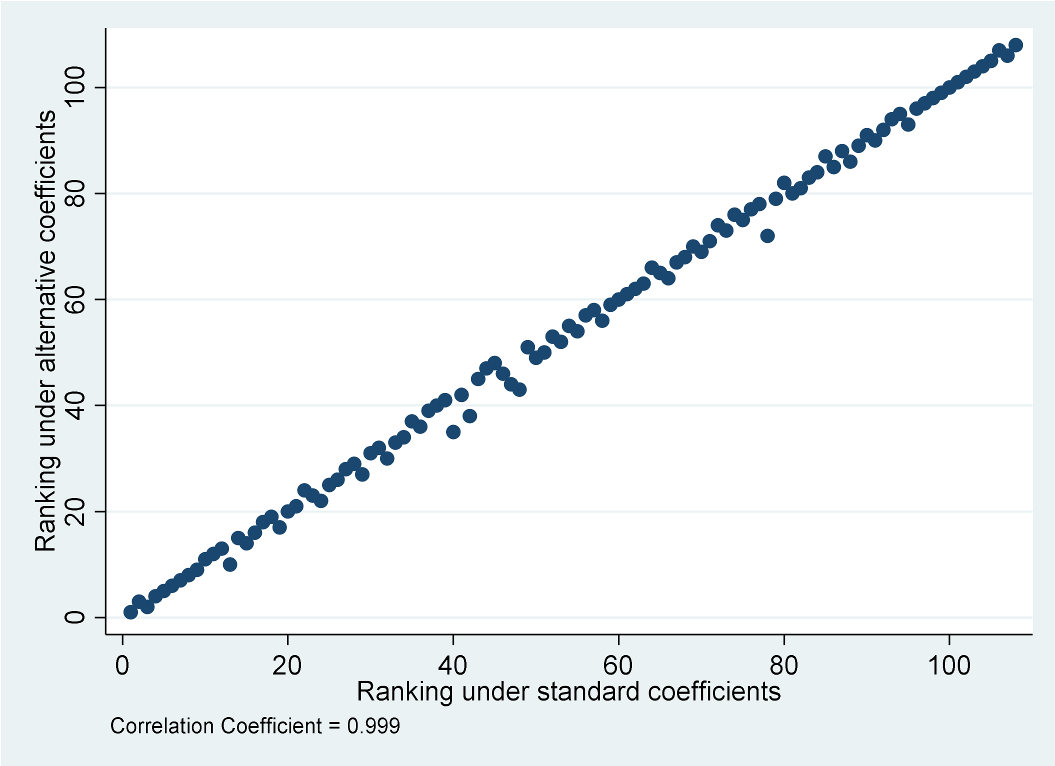

Figure 14 - Correlation between HCI rankings under two different coefficients setting. Reference year: 2016.

As it is easy to see, the observations lay almost perfectly along the 45° bysector, indicating that the ranking remains substantially unchanged under alternative assumption about the returns to health and education. The correlation coefficient between the two rankings is equal to 0.999.

INVALSI provides test scores calculated according to three different methods. The test scores are reported in their raw version, normalized and computed according to a Rasch model (Boone 2016). In the estimation, we have utilized the normalized test scores. We now test the robustness of the results with respect to different ways of computing the test scores. We re-compute the index using both the raw and the Rasch scores. As shown in figure 15, the results obtained using the two alternative ways of computing the test scores are almost identical (for a representation of the HCI using normalized test scores see figure 4Figure). Although the results of the Rasch model are slightly more different from the raw ones, the gap is still very small. Therefore, the different ways of computing the test scores do not appear to affect the results.

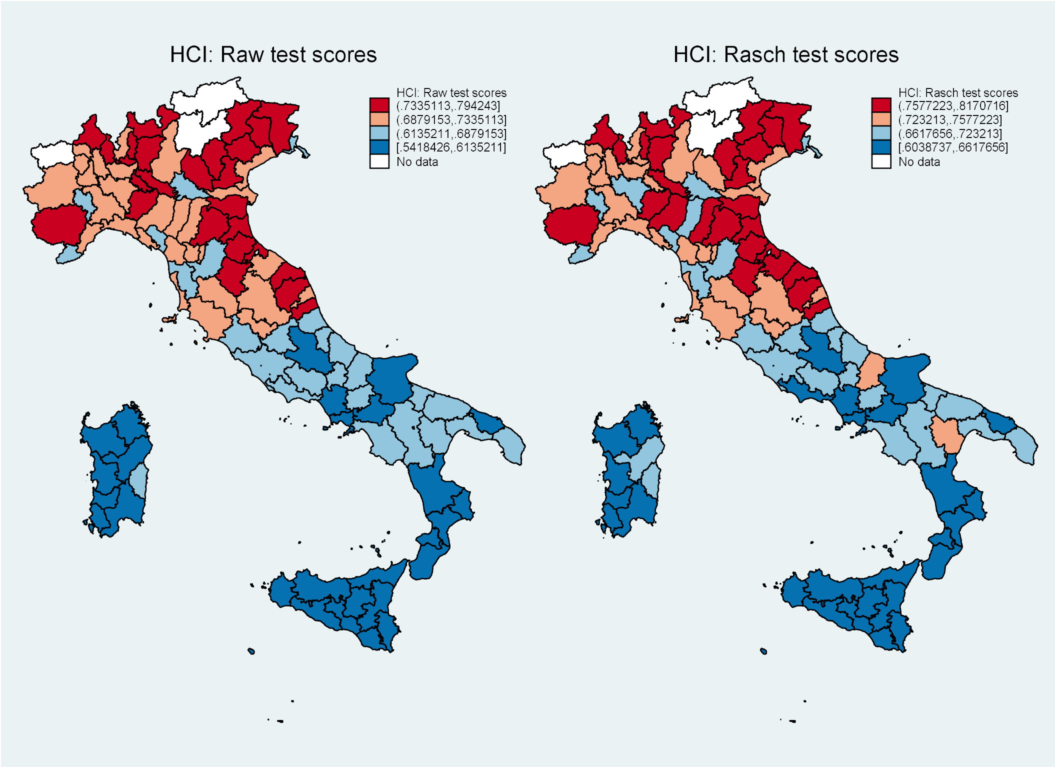

Figure 15 - Robustness check with respect to different test scores measure. On the left hand-side, raw test scores were used. On the right hand-side, Rasch test scores were used. Reference year: 2016.

Following Kraay (2018), we built the HCI using the net enrollment rate by Province and considering the test scores of children in the expected grade. In figure 16, we compare the HCI obtained using the net enrollment rate and the test score of children enrolled in the expected grade with those obtained using the gross enrollment rate and the test score of all children enrolled in the grade.

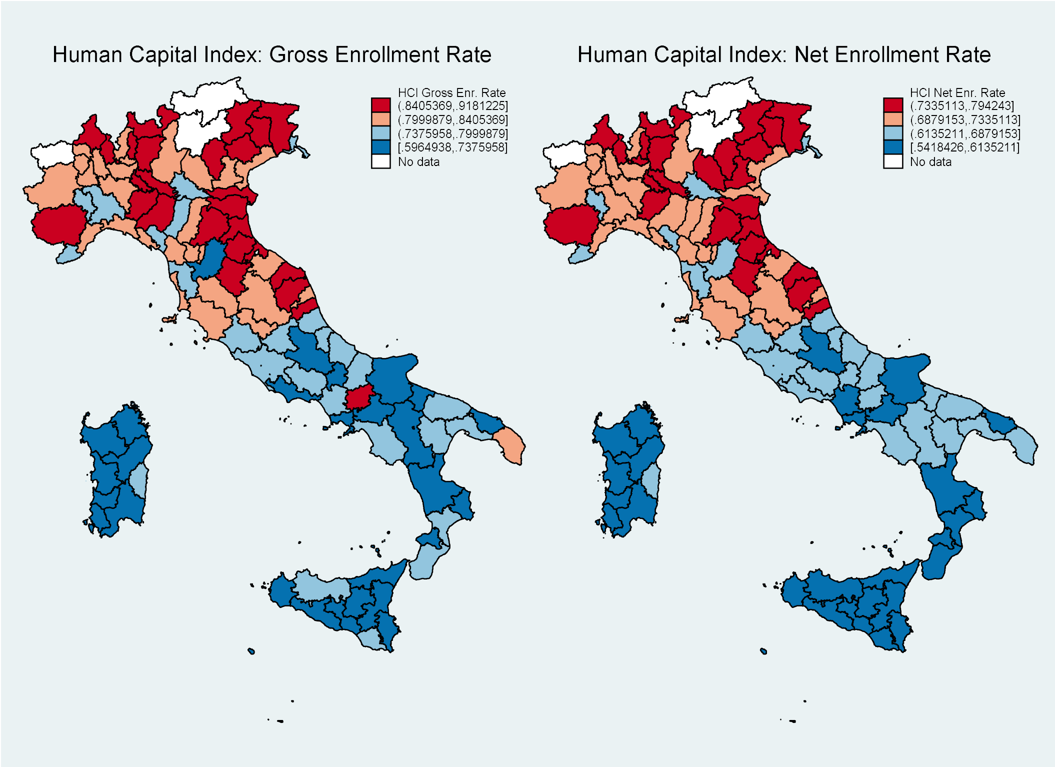

Figure 16 - Robustness check with respect to different enrolment rate measures. On the left hand side, it was used the gross enrolment rate. On the right hand side, it was used the net enrolment rate. Reference year: 2016.

As it is possible to see, the values of the HCI are obviously higher when the gross enrollment rate is employed. Nonetheless, the relative rankings are similar, suggesting the use of the net rather than the gross enrollment rate is not particularly relevant in the determination of the HCI.

5. Conclusions

We have developed a Human Capital Index for the Italian provinces following the methodology proposed by the World Bank. As partly expected, given the wide differences in socio economic conditions, the human capital with which youth enter the labour market in Italy show substantial variation across the country. From a comparison with World's countries, the HCI of Lecco, the Italian best performing Province, set itself between the two best performing world countries: South Korea and Singapore. The HCI of Vibo Valentia, the Italian worst performing Province according to the harmonized scores, is approximately equal to that of Jordan, ranked 79th among world countries. The data on the stock of human capital of youth confirm the permanence of substantially large differences across Italy. The dualism of the Italian economy still persists and, given the role played by human capital in determining growth, appears to be likely to continue.

The analysis, in fact, suggests only a slow convergence across provinces of the HCI between 2016 and 2017. The Human Capital Index is highly correlated with the average income per capita of the Provinces and with a set of socio economic characteristics of the provinces confirming the existence of structural differences across the Italian territory. Addressing the differences of the human capital stock of youth appears, therefore, a particularly important policy target to deal with the dualism of the Italian economy.

The use of World Bank Human Capital Index has allowed us to compare the ranking across Italian Provinces with World Bank's world ranking of countries. Nonetheless, when developing such an index, disaggregated at territorial level, for a high-income country as Italy, several considerations apply that need to be kept in mind and that could lead potentially to a refinement of the index.

In the case of high-income countries, a relative large share of youth continues on to tertiary education. There is evidence that tertiary education has high returns in terms of productivity. Therefore, not including this dimension in the index may be misleading.

As far as health is concerned, in high-income countries, malnutrition and life expectancy in adulthood might not fully reflect the stock of expected health of the young individual. Differences in life expectancy might in fact not fully reflect the expected productive capabilities of the individual. Information on the ability of the health system to limit morbidity may, for example, be more relevant.

The disaggregation at sub-national level poses additional problem linked to the internal mobility of individual across education institutions and local labour markets. While the former might not pose severe problems at the level of secondary education, the latter is potentially more relevant as a relatively large number of youth migrate within the country to enter the labour market (possibly after attending a tertiary level institution). The HCI, therefore, in this context has to be interpreted mostly as an indicator of the human capital provided locally to secondary education students rather than as a measure of the human capital with which youths enter the labour market in a specific area of the country.

Notwithstanding these limitations, that we hope to address in future work, the HCI at provincial level does offer a set of very relevant policy information that can be used to design intervention policies.

References

Blane, D. (1995). Social determinants of health-socioeconomic status, social class, and ethnicity. *American Journal of Public Health, 85*(7), 903-905.

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When and How? *Life Sciences Education, 15*(4), 1-7.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives, 27*(3), 79-102.

Filmer, D. P., Rogers, F. H., Angrist, N., & Sabarwal, S. (2018). Learning-Adjusted Years of Schooling (LAYS): Defining A New Macro Measure of Education. *Policy Research Working Paper Series*(8591), 1-44.

Folloni, G., & Vittadini, G. (2010). Human Capital Measurement: A Survey. *Journal of Economic Surveys, 24*(2), 248-279.

Istat. (2019). *Le Differenze Territoriali di Benessere: una lettura a livello provinciale.* Roma: Territori, Lettura di Statistiche.

Kraay, A. (2018). Methodology for a World Bank Human Capital Index. *Policy Research Working Paper Series*(8593), 1-57.

Patrinos, H. A., & Angrist, N. (2018). Global Dataset on Education Quality: A Review and Update (2000-2017). *Policy Research Working Paper Series*(8592), 1-35.

Quintano, C., Castellano, R., & Longobardi, S. (2009). A fuzzy clustering approach to improve the accuracy of italian student data: An experimental procedure to correct the impact of outliers on assessment test scores. *Statistica Applicata, 7*(2), 149-171.

Smith, A. (1723-1790). *An Inquiry into the Nature and Causes of the Wealth of Nations* (Vol. II). New York: Modern Library.

Weil, D. N. (2007). Accounting for the Effect of Health on Economic Growth. *The Quarterly Journal of Economics, 122*(3), 1265-1306.

Appendix

A. Human Capital Index by Province

RANK	PROVINCE	MACRO-AREA	REGION	HCI
1	LECCO	LOMBARDIA	NORTH WEST	0.79
2	FORLÌ-CESENA	EMILIA ROMAGNA	NORTH EAST	0.79
3	SONDRIO	LOMBARDIA	NORTH WEST	0.79
4	PIACENZA	EMILIA ROMAGNA	NORTH EAST	0.78
5	UDINE	FRIULI-VENEZIA G.	NORTH EAST	0.77
6	ASCOLI PICENO	MARCHE	CENTRE	0.77
7	BELLUNO	VENETO	NORTH EAST	0.77
8	ANCONA	MARCHE	CENTRE	0.77
9	VARESE	LOMBARDIA	NORTH WEST	0.76
10	BERGAMO	LOMBARDIA	NORTH WEST	0.76
11	PORDENONE	FRIULI-VENEZIA G.	NORTH EAST	0.76
12	VICENZA	VENETO	NORTH EAST	0.76
13	CUNEO	PIEMONTE	NORTH WEST	0.76
14	RIMINI	EMILIA ROMAGNA	NORTH EAST	0.76
15	TREVISO	VENETO	NORTH EAST	0.76
16	MACERATA	MARCHE	CENTRE	0.75
17	BOLOGNA	EMILIA ROMAGNA	NORTH EAST	0.75
18	CREMONA	LOMBARDIA	NORTH WEST	0.74
19	FERRARA	EMILIA ROMAGNA	NORTH EAST	0.74
20	RAVENNA	EMILIA ROMAGNA	NORTH EAST	0.74
21	LODI	LOMBARDIA	NORTH WEST	0.74
22	MONZA E DELLA BRIANZA	LOMBARDIA	NORTH WEST	0.74
23	VERONA	VENETO	NORTH EAST	0.73
24	VERBANO-CUSIO-OSSOLA	PIEMONTE	NORTH WEST	0.73
25	PADOVA	VENETO	NORTH EAST	0.73
26	AREZZO	TOSCANA	CENTRE	0.73
27	MODENA	EMILIA ROMAGNA	NORTH EAST	0.73
28	PESARO E URBINO	MARCHE	CENTRE	0.73
29	PARMA	EMILIA ROMAGNA	NORTH EAST	0.73
30	BRESCIA	LOMBARDIA	NORTH WEST	0.72
31	COMO	LOMBARDIA	NORTH WEST	0.72
32	ROVIGO	VENETO	NORTH EAST	0.72
33	VENEZIA	VENETO	NORTH EAST	0.72
34	PERUGIA	UMBRIA	CENTRE	0.72
35	SAVONA	LIGURIA	NORTH WEST	0.71
36	BIELLA	PIEMONTE	NORTH WEST	0.70
37	TORINO	PIEMONTE	NORTH WEST	0.70
38	FERMO	MARCHE	CENTRE	0.70

39	NOVARA	PIEMONTE	NORTH WEST	0.70
40	VERCELLI	PIEMONTE	NORTH WEST	0.70
41	SIENA	TOSCANA	CENTRE	0.70
42	GENOVA	LIGURIA	NORTH WEST	0.70
43	PISTOIA	TOSCANA	CENTRE	0.70
44	LA SPEZIA	LIGURIA	NORTH WEST	0.70
45	MILANO	LOMBARDIA	NORTH WEST	0.70
46	TERNI	UMBRIA	CENTRE	0.70
47	LUCCA	TOSCANA	CENTRE	0.70
48	GROSSETO	TOSCANA	CENTRE	0.70
49	PRATO	TOSCANA	CENTRE	0.70
50	PAVIA	LOMBARDIA	NORTH WEST	0.69
51	ALESSANDRIA	PIEMONTE	NORTH WEST	0.69
52	REGGIO NELL'EMILIA	EMILIA ROMAGNA	NORTH EAST	0.69
53	LIVORNO	TOSCANA	CENTRE	0.69
54	MANTOVA	LOMBARDIA	NORTH WEST	0.69
55	TRIESTE	FRIULI-VENEZIA G.	NORTH EAST	0.68
56	LECCE	PUGLIA	SOUTH	0.68
57	PISA	TOSCANA	CENTRE	0.68
58	GORIZIA	FRIULI-VENEZIA G.	NORTH EAST	0.68
59	IMPERIA	LIGURIA	NORTH WEST	0.67
60	PESCARA	ABRUZZO	SOUTH	0.67
61	CHIETI	ABRUZZO	SOUTH	0.67
62	CAMPOBASSO	MOLISE	SOUTH	0.67
63	BENEVENTO	CAMPANIA	SOUTH	0.67
64	MATERA	BASILICATA	SOUTH	0.67
65	ITALIA	ITALIA	ITALIA	0.66
66	ASTI	PIEMONTE	NORTH WEST	0.66
67	VITERBO	LAZIO	CENTRE	0.66
68	MASSA-CARRARA	TOSCANA	CENTRE	0.66
69	FIRENZE	TOSCANA	CENTRE	0.65
70	TERAMO	ABRUZZO	SOUTH	0.65
71	ROMA	LAZIO	CENTRE	0.64
72	BARI	PUGLIA	SOUTH	0.64
73	RIETI	LAZIO	CENTRE	0.64
74	FROSINONE	LAZIO	CENTRE	0.63
75	SALERNO	CAMPANIA	SOUTH	0.63
76	BARLETTA-ANDRIA-TRANI	PUGLIA	SOUTH	0.63
77	TARANTO	PUGLIA	SOUTH	0.63
78	OGLIASTRA	SARDEGNA	ISLANDS	0.63
79	ISERNIA	MOLISE	SOUTH	0.62
80	POTENZA	BASILICATA	SOUTH	0.61
81	LATINA	LAZIO	CENTRE	0.61
82	REGGIO DI CALABRIA	CALABRIA	SOUTH	0.61
83	RAGUSA	SICILIA	ISLANDS	0.61
84	NUORO	SARDEGNA	ISLANDS	0.61
85	FOGGIA	PUGLIA	SOUTH	0.61
86	L'AQUILA	ABRUZZO	SOUTH	0.60
87	CATANZARO	CALABRIA	SOUTH	0.60
88	ENNA	SICILIA	ISLANDS	0.60
89	BRINDISI	PUGLIA	SOUTH	0.60

90	CALTANISSETTA	SICILIA	ISLANDS	0.60
91	CASERTA	CAMPANIA	SOUTH	0.60
92	CAGLIARI	SARDEGNA	ISLANDS	0.59
93	PALERMO	SICILIA	ISLANDS	0.59
94	AGRIGENTO	SICILIA	ISLANDS	0.59
95	AVELLINO	CAMPANIA	SOUTH	0.59
96	NAPOLI	CAMPANIA	SOUTH	0.58
97	SASSARI	SARDEGNA	ISLANDS	0.58
98	COSENZA	CALABRIA	SOUTH	0.58
99	VIBO VALENTIA	CALABRIA	SOUTH	0.57
100	ORISTANO	SARDEGNA	ISLANDS	0.57
101	SIRACUSA	SICILIA	ISLANDS	0.57
102	TRAPANI	SICILIA	ISLANDS	0.57
103	CATANIA	SICILIA	ISLANDS	0.57
104	MESSINA	SICILIA	ISLANDS	0.56
105	CROTONE	CALABRIA	SOUTH	0.56
106	OLBIA-TEMPIO	SARDEGNA	ISLANDS	0.56
107	CARBONIA-IGLESIAS	SARDEGNA	ISLANDS	0.55
108	MEDIO CAMPIDANO	SARDEGNA	ISLANDS	0.54

B. Human Capital Index Distribution

In figure B.1., we show the distribution of the Human Capital Index. From the histogram, it is possible to see that the distribution has three main modal points. The first one corresponding to a value of the human capital index slightly lower than 0.6, the second one corresponding to a value of the human capital index of 0.7 and the third one corresponding to a value of the human capital index of approximately 0.75.

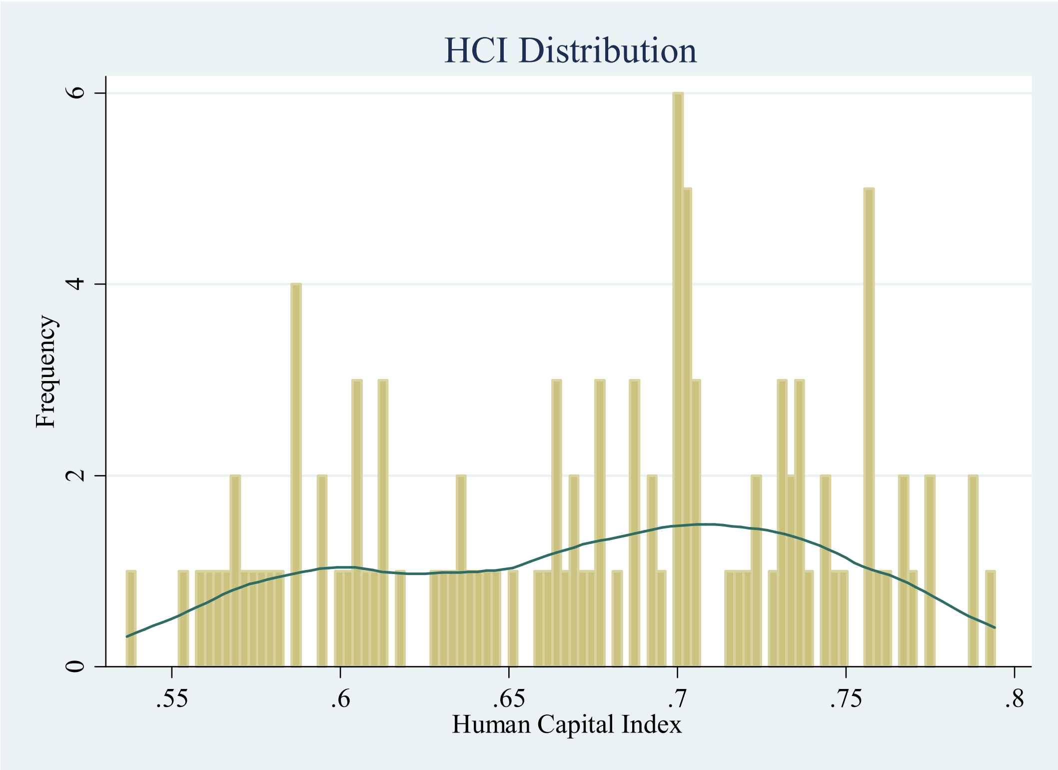

Figure B.1. - Distribution of the Human Capital Index values. Reference year: 2016.

C.Harmonization of the test scores

The ratio between the observed test scores and the conversion factor returns the test score values that we would observe using TIMSS and PISA data. More in detail, to obtain the harmonized test score, we normalize the observed test scores to have an average of 500 and a standard deviation of 100 (as the PISA and TIMSS test scores). To obtain the harmonized test scores, we divide the normalized value by the following conversion factor:

$$CONV_{INV-PISA}^{II} = \sum_{t=2011/12,2014/15} \left(\frac{\sum_{g=1}^{5} s_{g,INV,t}^{II} /5}{\sum_{g=1}^{5} s_{g,PISA,t}^{II} /5} \right) /2$$

$$CONV_{INV-TIMSS}^{II} = \sum_{t=2011/12,2014/15} \left(\frac{\sum_{g=1}^{5} s_{g,INV,t}^{II} /5}{\sum_{g=1}^{5} s_{g,TIMSS,t}^{II} /5} \right) /2$$

where the first conversion factor is used to convert INVALSI test scores in reading to PISA's while the second is used to convert INVALSI test scores in mathematics to TIMSS's. The term $s_{g,j,t}^{i}$ is the average test score observed for the geographical area g, at time t, using data j referred to school level i (where i identifies primary or secondary school). In the world comparison, we built the final expected LAYS using Singapore (rather than Lecco) as a benchmark. Following these adjustments, we still have some small differences between the HCI for Italy computed by the World Bank and that computed by us. This most likely due to marginal differences in the enrollment rates used in the estimations. To enhance comparability we have, therefore, rebased our index to coincide with that of the World Bank at national level.