

A COMPARATIVE ANALYSIS OF LITERACY RATE IN CONTRIBUTING TO SOCIAL EXCLUSION INSIGHTS

Edgardo Bucciarelli* - Fabrizio Muratore† - Iacopo Odoardi† - Carmen Pagliari*

Abstract

Our contribution aims to analyze the complex relationship between the phenomena of social exclusion and literacy, moreover considering the significant implication on economic growth. The aim, by analyzing cross-country, is first to describe the situation of social exclusion with the use of specific socio-economic variables, and second to compare the levels of education and training of each country. These two phenomena are mutually influenced, as a low level of literacy in affecting the employment status precludes the possibility to enter and operate freely in society, while poverty and persistent social exclusion of a person or family makes difficult to address appropriate educational and training paths. Our analysis has therefore rejoined two issues which are now particularly examined and influence almost all decisions made by policy makers, especially in the Western world. The first issue is the level of education, and through appropriate investment it should constitute the human capital of a country, the second is the relational conditions of society, which underlies the so-called social capital. Together these two types of intangible capitals constitute a strong support for the long-term development of a nation. Our quantitative analysis is also addressed to detect differences and peculiarities between the different national realities, with the ultimate aim to understand which socio-economic variables affect the processes of education more directly.

JEL classification: C82; I21; O50; O57; Y10.

Keywords: social exclusion, literacy rate, maximum likelihood, hierarchical cluster.

* Department of Quantitative Methods and Economic Theory, “*Gabriele d’Annunzio*” University of Chieti-Pescara, Viale Pindaro n. 42, Pescara 65127, Italy, Tel.: +39 085 453 7585 and +39 085 453 7980; correspondence may be address to: e.bucciarelli@unich.it

† School of Advanced Studies “*Gabriele d’Annunzio*”, Doctorate of Philosophy, “*Gabriele d’Annunzio*” University of Chieti-Pescara, Viale Pindaro n. 42, Pescara 65127, Italy, Tel.: +39 085 453 7679.

Responsibility for the contents of this article is entirely ours and it should not be attributed to our affiliated institutions. We thank Giuliana Parodi, Dario Sciulli and participants at the XXV National Conference of Labour Economics, AIEL, Pescara (Italy), September 9-10, 2010 for suggestions, useful comments and support.

1. Introduction

The level of education for an individual represents, especially in modern Western societies, a potential effective indicator of the level of working capacity, the so-called productivity, but also implies other less observable skills, but equally significant. These are the essential knowledge for living in society as dynamic players, without incurring the risk of being excluded from those who are its traditional and contemporary activities. But today there are many limitations to the normal training of an individual, and in some cases, to whole groups of disadvantaged people. Considering for example the problem of poverty, a phenomenon that often affects the most vulnerable groups, including their children, the elderly and ethnic minorities. Children living in families at social risk or below the poverty line, see foreclosed the best channels of education and training, finding obstacles in the first few years of training school, and this could affect the subsequent paths of training and prevent achieve adequate levels of education. For adequate levels of education we mean the level of training which, for each culture and society, is considered optimal in order to access to stable employment and profitable, so as to have no risk of not being able to participate in society. However, if households living below the poverty line are the most at risk, we must remember those with only one parent, those who live in neighborhoods with little social organization, and even those belonging to ethnic or religious minority. Moreover, strong correlation between the terms Illiteracy and social exclusion comes to us in some more complete definitions of literacy have been given over time. An example is provided by the General Conference of UNESCO in 1978 which provides several definitions of literacy. The first: *“A person is literate who can with understanding both read and write a short simple statement on his everyday life.”*, while the second is more complex: *“A person is functionally literate who can engage in all those activities in which literacy is required for effective functioning of his group and community and also for enabling him to continue to use reading, writing and calculation for his own and the community’s development.”* They range from the simple ability in writing and reading to the reasoning skills and development of different ability and expertise, to include situations about social life and the use of their knowledge adequately. We are not referring merely to the increasing of more cognitive abilities which allow to reach higher levels of understanding, but also the manner in which those who have these skills can exploit them to coexist in a stable and balanced society to which they belong to. This particularly includes the definition offered by the International Adult Literacy Survey (see OECD/Statistics Canada, 2000): *“The ability to understand and employ printed information in daily activities, at home, at work and in the community - to achieve one’s goals, and to develop one’s knowledge and potential.”* Definition used by the OECD (2010) which adds that in more *“[...] Differences in levels of literacy matter both economically and socially: literacy affects, inter alia, labour flexibility and quality, employment, training opportunities, income from work and wider participation in civic society.”* Furthermore, the IALS Final Report (OECD / Statistics Canada, 2000) shows a number of relationships about literacy: for example, those countries with higher literacy scores had higher labour force participation and shorter work hours. Countries which have a high proportion of adults with low prose skills, and (conversely) those which have a low proportion with high prose skills, had lower GDP per capita. The higher the proportion of adults with low prose skills, and the lower the proportion with high prose skills, the lower that country’s GDP per capita.

Public intervention may partially cope and withstand the mentioned problems related to lack of literacy, with a significant long-term goals in order to obtain a plentiful and robust level of national human capital. This is now a fundamental part of many economic studies concerning growth and development, as it is considered, when it is registered in a high quality, safe tool to foster the virtuous processes of economic growth. But the complex phenomenon of education and training of individuals, cannot be regarded only as general educational teaching base of general educational foundations, for both the need for a training program more complete and durable (just think of the

increasingly popular process called lifelong learning) and the standpoint that we could define quality. This means not only prepared trainers, and adequate public and private investment, but also a rational project to provide all individuals involved in education in the sense of common rules and social (see, for example, Grossman and Kim, 1997). These are the foundations for building a solid social capital, itself a determinant of socio-economic development processes (see among others Gradstein and Justman, 2002, who have combined in the relationship between social capital and economic growth, also the education processes).

People who cannot, by choice or compulsion, reach satisfactory levels of literacy, are to be among the weakest in society, and go toward the phenomena of social exclusion. It can be observed the problems caused by poor literacy through two joint issues. The first concerns the limited prospects of finding employment opportunities for safe and protected, the second considers the defective contribution which people supplies in the socio-economic system, above all in terms of labour productivity. The difficulty in finding a job involves a certain separation from society, depriving the individual of the ability to fully exploit the possibilities offered by his contemporary world. The individuals, though not in poverty, will have the opportunity to engage in specified conduct that are characteristic of each society, and constitute a general need to be met. Therefore, the path that is likely to take is degenerative, and leads to conditions close to poverty as relative, which tends to become a phenomenon that is passed between generations. Poor parents in fact, cannot guarantee the optimal education for their children (about the ratio of costs for education, inequality and transition between generations, see among others Grossman, 2008; and Galor and Moav, 2004), while parents who are socially excluded represent a source of social exclusion for their children. The education level of a person can be found through the years of school attended, or through qualifications acquired. In particular, in this work we observe the influence which some typical variables, related to the processes of social exclusion, can have on the average level of education of a country. It is known that the economic conditions may affect the ability to achieve high average levels of education, but also the strength of education and social capital, based on it, constitute a lever of development (see among others Temple and Johnson, 1998). Gradstein and Justman (2002) analyze the importance of a broad common cultural basis to start those basic functions of effective interaction between individuals. The spread of education and in general of literacy may be a fundamental policy goal in trying to establish virtuous processes from an economic standpoint. Essentially, the main aim of our empirical study is to demonstrate how certain variables¹ which characterize social exclusion have obvious influences on the literacy rate. Then it is briefly compared the effect of this variable on the level of economic development of a country.

The framework of the paper is organized as follows. We start in section 2 showing the methodology used. Then we proceed in section 3 with a preliminary empirical investigation on literacy rate and the expectancy variables. In section 4 are presented the results of the multivariate regression model. The findings of the factor analysis with maximum likelihood method and VARIMAX rotation are dealt with in section 5. Furthermore, section 6 focuses on multivariate regression model for literacy rate with maximum likelihood components. We conclude in section 7.

¹ We have collected the cross-country dataset from UNESCO (2010) and World Bank (2010). For the cross-country analysis we provide see also Levine and Renelt (1991, 1992), de Gregorio and Lee (2002), Hoover and Perez (2004).

2. Methodology applied for empirical analysis on literacy rate for 30 countries

In this paper we analyze literacy rate by least squares method, factor analysis with maximum likelihood method and hierarchical cluster. The method of least squares is a standard approach to the approximate solution of over determined systems (Moser, 1996; Freund 2003), i.e. sets of equations in which there are more equations than unknowns. Least squares means that the overall solution minimizes the sum of the squares of the errors made in solving every single equation. The most important application is in data fitting: the best fit in the least-squares sense minimizes the sum of squared residuals, a residual being the difference between an observed value and the value provided by a model. We consider a linear regression model: hence the model comprises a linear combination of the parameters:

$$f(x_i, \beta) = \sum_{j=1}^m \beta_j \phi_j(x_i) \quad (1)$$

where the coefficients, ϕ_j , are functions of x_i . Letting:

$$X_{ij} = \frac{\partial f(x_i, \beta)}{\partial \beta_j} = \phi_j(x_i) \quad (2)$$

Then we can see in case (2) the least square estimate (or estimator, in the context of a random sample), β is given by:

$$\beta = (X^T X)^{-1} X^T y \quad (3)$$

Hence we consider the response variable as a linear function of the regressors:

$$y_t = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon_t \quad (4)$$

In our paper we analyze literacy rate (LR) in function of school enrollment consisting in pre-primary (PRE), primary (PRY), secondary (SEC) and tertiary (TER); GDP real growth rate (GDP); long unemployment rate (LUR); public spending on education as percentage of GDP (PSE) and public spending on education as percentage of government expenditure (PGE); children out of school (COS). The analysis regard 30 countries by different continents and we consider mostly countries of Organizations for Economic Co-operation and Development (OECD). The model we provide is the following:

$$LR = \beta_0 + \beta_1 PRE + \beta_2 PRY + \beta_3 SEC + \beta_4 TER + \beta_5 GDP + \beta_6 LUR + \beta_7 PSE + \beta_8 PGE + \beta_9 COS \quad (5)$$

Therefore in the next study we apply factor analysis in order to reduce the high number of variables of the model (5). Factor analysis is a statistical method used to describe variability among observed variables in terms of a potentially lower number of unobserved variables called factors (Mardia, 1980). In other words, it is possible, for example, that variations in three or four observed variables mainly reflect the variations in a single unobserved variable, or in a reduced number of unobserved variables. Factor analysis searches for such joint variations in response to unobserved

latent variables. The observed variables are modeled as linear combinations of the potential factors, plus an error terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset. Factor analysis is related to principal component analysis (PCA) but it is not identical, because PCA performs a variance-maximizing rotation of the variable space and it takes into account all variability in the variables. In contrast, factor analysis estimates how much of the variability is due to common factors (communality). The two methods become essentially equivalent when the error terms in the factor analysis model (the variability not explained by common factors, see below) can assume the constant variance. In fact principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much as possible of the variability in the data, and each succeeding component accounts for as much as possible of the remaining variability. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA is theoretically the optimum transform for given data in least square terms. For a data matrix, X^T , with zero empirical mean (the empirical mean of the distribution has been subtracted from the data set), where each row represents a different repetition of the experiment, and each column gives the results from a particular problem, the PCA transformation is given by:

$$Y^T = X^T W = V \Sigma^T \quad (6)$$

where the matrix Σ is an m-by-n diagonal matrix with non-negative real numbers on the diagonal and $W \Sigma^T V$ is the singular value decomposition (svd) of X . In our analysis we tried to apply principal component analysis but we find not-good results. Therefore we use maximum likelihood method with VARIMAX rotation considering an extraction based on eigenvalue with eigenvalue greater than 0,5. Indeed, in this kind of choice we find a good response of components. Before trying this method, we analyze literacy rate also considering an extraction base on eigenvalue greater than 1 but we discover a low result. In fact maximum likelihood estimation (MLE) is a popular statistical method used for fitting a statistical model to data, and providing estimates for the model's parameters (Besset, 2001). The method of maximum likelihood corresponds to many well-known estimation methods in statistics. The sample mean is then the maximum likelihood estimator of the population mean, and the sample variance is a close approximation to the maximum likelihood estimator of the population variance. For a fixed set of data and underlying probability model, the method of maximum likelihood selects values of the model parameters which maximize the likelihood function. Maximum likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems.

In applying MLE we suppose that there is a sample x_1, x_2, \dots, x_n of n independent and identically distributed (i.i.d.) observations, coming from an unknown distribution $f_0(\cdot)$. It is however known that the function f_0 belongs to a certain family of distributions $\{f(\cdot | \theta), \theta \in \Theta\}$, called the parametric model, so that $f_0 = f(\cdot | \theta_0)$. The value θ_0 is unknown and is referred to as the “true value” of the parameter. It is desirable to find some $\hat{\theta}$ (the *estimator*) which would be as close to the true value θ_0 as possible. Both the observed variables x_i and the parameter θ can be vectors. The variables x_i may be non-i.i.d., in which case the formula below for joint density will not separate into individual terms; however the general principles would still apply. To use the method of maximum likelihood, one first specifies the joint density function for all observations. For i.i.d. sample this joint density function will be

$$f(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta) \cdot f(x_2 | \theta) \cdots f(x_n | \theta) \quad (7)$$

We may extend the domain of the density function so that the density is also a function of the parameter θ . Then, for a given sample of data with observed values x_1, x_2, \dots, x_n , the extended density can be considered a function of the parameter θ . This extended density is the likelihood function of the parameter:

$$\mathcal{L}(\theta | x_1, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad (8)$$

However, in general, the likelihood function is not a probability density. In fact, it does not need to be an additive function, thus it is not a probability measure. In practice it is often more convenient to work with the logarithm of the likelihood function, $\ln \mathcal{L}$, called the log-likelihood, or its scaled version, called the average log-likelihood:

$$\ln \mathcal{L}(\theta | x_1, \dots, x_n) = \sum_{i=1}^n \ln f(x_i | \theta), \quad \hat{\ell} = \frac{1}{n} \ln \mathcal{L} \quad (9)$$

Indeed, $\hat{\ell}$ estimates the expected log-likelihood of a single observation in the model. The method of maximum likelihood estimates θ_0 by finding a value of θ that maximizes $\hat{\ell}(\theta | x)$. This method of estimation is a maximum likelihood estimator (MLE) of θ_0 :

$$\hat{\theta}_{mle} = \arg \max_{\theta \in \Theta} \hat{\ell}(\theta | x_1, \dots, x_n) \quad (10)$$

In applying maximum likelihood estimation, we can identify a point estimate referred to each countries considered in our analysis. Furthermore in the application by maximum likelihood we use VARIMAX rotation (Jakeman, 2008). We know that in quantitative methods, a VARIMAX rotation is a change of coordinates used in principal component analysis and factor analysis which maximizes the sum of the variances of the squared loadings. That is, it seeks a basis that most economically represents each individual, so that each country can be well described by a linear combination of only a few basic functions:

$$R_{VARIMAX} = \arg \max_{\mathbf{R}} \left(\sum_{j=1}^k \sum_{i=1}^p (\Lambda \mathcal{R})_{ij}^4 - \frac{\gamma}{p} \sum_{j=1}^k \left(\sum_{i=1}^p (\Lambda \mathcal{R})_{ij}^2 \right)^2 \right) \quad (11)$$

where $\gamma = 1$ for VARIMAX. Variance maximizing rotation is often used in surveys to see how groups of countries measure the same phenomenon. In our case how the 30 countries considered distance themselves.

Ultimately we use hierarchical clustering (Mauler, 2001), which is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types, but we use only agglomerative method that is a bottom up approach, in fact each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. In general, the merges and splits are determined in a greedy manner. The results of hierarchical

clustering are usually presented in a dendrogram. In order to decide which clusters should be combined (for agglomerative) a measure of dissimilarity between sets of observations is required. In most methods of hierarchical clustering, this is achieved by use of an appropriate metric (a measure of distance between pairs of observations), and a linkage criteria which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets. We chose Euclidean distance:

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \quad (12)$$

We used the linkage criteria maximum or complete linkage clustering that determines the distance between sets of observations as a function of the pairwise distances between observations:

$$\max\{d(a, b) : a \in A, b \in B\} \quad (13)$$

We apply this method of hierarchical cluster in order to identify any consistency between the countries considered, and part of the same cluster as well as the heterogeneity between clusters.

3. Preliminary empirical investigation on literacy rate and the expectancy variables

The following analysis refers to 30 countries some of which are OECD members, ranging a period of time 2007/2009. The first analysis performs the calculation of average values, lowest and highest values reported for each variable under study. Variables that we analyze are: literacy rate, GDP real growth rate, pre-primary, primary, secondary and tertiary instructions as percentage of total population by every instruction of every country, long term unemployment rate, public spending as percentage GDP and public spending as percentage of government expenditure, children out of school as percentage of total children out of school in vary countries. In the next page we show the results obtained from descriptive analysis:

indicator	mean	lowest	highest	indicator	mean	lowest	highest
literacy rate	98,254	88,7	99,8	tertiary	0,413	0,01	0,968
		Turkey	Latvia			Netherlands	U.S.
GDP real growth rate	0,003	-0,041	0,045	long unemployment rate	1,760	0,045	5,988
		Latvia	Poland			Mexico	Slovakia
pre-primary	0,891	0,159	1,233	public expen. perc. GDP	0,051	0,019	0,086
		Turkey	Spain			Austria	Japan
primary	1,058	0,549	1,577	public expen. perc. Gov.	0,132	0,034	0,281
		Portugal	Spain			Portugal	Latvia
secondary	0,785	0,183	1,479	children out of school	0,030	0,0002	0,416
		Netherlands	Australia			Luxembourg	U.S.

Table 1: average, lowest and highest values for dependent and expectancy variables.

Literacy rate has a mean value equal to 98,254%, this average is referred to 30 countries considered in our analysis. The minimum is 88, 7% and is represented by Turkey, the maximum value is represented by Latvia and is almost 100%. GDP real growth rate has its minimum value in Latvia, the same regards other countries such as Hungary and Ireland which show negative values, while the highest refers to Poland. The average GDP real growth rate is near to zero, because in recent years there has been a stagnation and recession of economy. The several education levels get the minimum value by Turkey, Portugal and Netherlands in the four categories considered. Slovakia shows the highest value reported for long unemployment rate. Furthermore, Austria and Portugal are the countries with lower usage rates of public spending in education. Finally the variable related to children out of school shows its maximum value in U.S. and this percentage is always conditioned by a high population abundance, and we note simultaneously a high degree of school drop-out. In the next section we begin the analysis of the literacy rate applying multivariate regression method, maximum likelihood method and clustering analysis to highlight the differences in education levels in recent years in the countries considered.

4. Results from the multivariate regression model

A first analysis of the variables related to education highlights a high variability in data due to different density of countries considered. For example, United States has only in the pre-primary school more than 12 million of people, whereas U.S. population exceeds 300 million inhabitants (source: data from the U.S. Census Bureau, 2010). This indicates that data is based also on country's

population size. The same is true for Mexico, Japan and France, which show a high abundance of individuals enrolled in different educational levels. In contrast, countries which show less members are represented by Iceland, Ireland, Latvia, Luxembourg and Slovenia. This variability is also confirmed by symmetry and kurtosis indices showing high values; this indicates an asymmetry of these distributions compared to normal one. We consider that parameter is normal when most of data are distributed in average value, however in our results values are distributed more even in areas of tail. Hence, literacy rate and children out of school show an asymmetry of information: it is apparent, as there are quite different from their percentages. Some countries have 99% of schooling, but many others are below 90%. While long term unemployment rate, public spending and GDP real growth rate report almost a normal distribution with low variability. Examining interdependence between variables we notice strong groupings of countries in relation to education levels and economic growth. Considering different levels of education relate to each other, we note high direct relations. However, we note that in most countries the literacy rate has percentages between 90 and 99%. While long unemployment rate shows a clear grouping of countries in the lowest percentages of literacy rate. This indicates that with increasing of the degree of education levels the long unemployment rate decrease, as with increasing grades of schooling individuals encounter greater job opportunities. Our aim here is to analyze literacy rate relationship with other variables presented in section 3. We use a multivariate regression analysis and mathematical model proposed is (4), while in terms of variables is (5). Before applying (5) we perform other types of relationships: we have first investigated literacy rate as a function only of education variables, then literacy rate only with variables related to social exclusion. Observing our results of statistical tests for significance of the parameters and considering the best fitting to data, we choose to consider model as a whole: indeed, in order to identify the optimal model applied we use also education variables as percentages of total education in different countries; resulting data are not satisfactory for the significance of the parameters in any kind of linear combination calculated. Thus, combination of economic growth, education levels and social exclusion variables leads to a better result than analytical models considered singly. Following we show the model results:

		adjusted R	std. error of
R	R square	square	the estimate
0,472	0,223	-0,081	0,0227069

Table 2: determination index.

constant	0,988 (0,024)	tertiary	-3,476E-5 (0,001)
GDP	-0,003 (0,247)	long unemployment rate	0,001 (0,003)
pre-primary	-0,836 (0,434)	public expen. perc. GDP	0,121 (0,274)
primary	0,090 (0,101)	public expen. perc. Gov.	0,086 (0,079)
secondary	-0,048 (0,070)	children out of school	-0,018 (0,054)

Table 3: multivariate analysis of literacy rate.

In table 3 we report parameters resulting for each variable analyzed and we report in brackets standard error for each parameter. The results reveal positive and negative coefficients. Parametric values obtained are very small and tend to zero for those coefficients concerning variables of social exclusion. The significance of the coefficients, finally, refers to pre-primary education whereas a margin of error of 0,05. Hence, the coefficient values are not significant for low margins of error. Only constant obtains the significance of the parameter for the analysis. We note also a low correlation and determination index stands on 0,223: this value does not give a normal adaptation to observed data.

The overall results are not satisfying: we have not found any significance with variables related to education or those linked with social exclusion. This result lead us to use other analysis, choosing to apply factor analysis. Indeed, we note that there is a redundancy of information due to the presence of variables that are quite similar to each other: i.e. in the case of education we have many variables. This redundancy of information is known as a multicollinearity of variables and leads to greater dispersion of values; therefore to eliminate it, we chose to use factor analysis, but in this case we apply different types of analysis. The first one refers to principal components using (6). Results are not quite satisfactory, in fact we find a high variability of components and we cannot explain the phenomenon of literacy rate. The subsequent analysis is to apply maximum likelihood method. In the next section we report results of that.

5. Factor analysis with maximum likelihood method and VARIMAX rotation

In the following factor analysis we apply the maximum Likelihood method to identify new components of the model (10). Therefore we use this method with VARIMAX rotation (11) considering an extraction based on eigenvalue with eigenvalue greater than 0,5. Below we show the results:

	Initial	Extraction
GDP	0,190	0,608
pre-primary	0,347	0,828
primary	0,237	0,731
secondary	0,107	0,728
tertiary	0,505	0,916
long unemployment rate	0,208	0,958
public expen. perc. GDP	0,250	0,935
public expen. perc Gov.	0,287	0,384
children out of school	0,489	0,610

Table 4: initial and extraction values with maximum likelihood method.

In table 4 we report the initial and extraction of each variable according to maximum likelihood method. We note that initial value of each variable increases with extraction. In particular there is a good extraction in variable referred to long unemployment rate and public spending as percentage of GDP. Once that we calculated extraction we try to reduce variability of the phenomenon and eliminate multicollinearity in the distribution with maximum likelihood method. Following we show results:

Factor	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
			Cumulative %			Cumulative %			Cumulative %
	Total	% of Variance		Total	% of Variance		Total	% of Variance	
1	2,247	22,467	22,467	1,440	14,401	14,401	1,229	12,295	12,295
2	1,595	15,952	38,420	1,578	15,778	30,179	1,050	10,500	22,795
3	1,344	13,441	51,861	0,888	8,875	39,054	0,973	9,729	32,524
4	1,173	11,733	63,594	1,058	10,575	49,629	0,956	9,562	42,086
5	0,878	8,785	72,379	0,890	8,896	58,525	0,915	9,151	51,237
6	0,830	8,297	80,676	0,585	5,852	64,377	0,860	8,604	59,841
7	0,707	7,074	87,750	0,488	4,879	69,257	0,754	7,541	67,382
8	0,561	5,614	93,365	0,492	4,918	74,175	0,679	6,792	74,175
9	0,410	4,099	97,464						
10	0,254	2,536	100,000						

Table 5: initial eigenvalues, extraction sums of squared and rotation sums of squared.

In table 5 we report initial eigenvalues, extraction as sum of the squares and rotation method used: this method still shows a great dispersion of the variability. In fact, the main components are a very high number. This result is expected because reference data set is represented by a wide variety of variables. In fact, with four components we find an explanation of variability by 63%. With eight components we obtain 93% of variability of the phenomenon. Once we calculate eigenvalues we report results of factor matrix:

	Factor							
	1	2	3	4	5	6	7	8
long unemployment rate	0,926	-0,278	0,144					
public expen. perc Gov.	-0,387	-0,151			0,186	0,244	0,158	0,297
public expen. perc GDP	0,184	0,858	0,404					
children out school		0,566	-0,339		0,143	0,205	0,164	0,269
tertiary	0,545	0,499	-0,556	0,219				
pre-primary		-0,248	0,306	0,779			-0,224	
secondary		0,162		0,139	0,703	-0,403		0,133
primary	-0,247		0,265	-0,301	0,450	0,507	-0,213	
GDP		-0,242	-0,150	-0,115	0,362		0,254	-0,555

Table 6: factor matrix.

In factor matrix we show the importance of each component on every variable analyzed. The first component explains in particular the phenomenon of long unemployment rate, while the second one relates to public spending as percentage of GDP. The fourth, fifth and sixth are related to the several education levels. Then applying VARIMAX rotation, we obtain different results:

	Factor							
	1	2	3	4	5	6	7	8
Tertiary	0,857	0,198	0,135	-0,292		-0,181		
Children out school	0,632	-0,135	0,173		-0,335		0,140	-0,139
Long unemployment		0,965		-0,116				
Public perc GDP	0,209		0,924					-0,165
Primary				0,839				0,104
Public perc Govern		-0,224	-0,220	0,375		0,329		-0,142
Pre primary	-0,145				0,850	0,262		
Secondary							0,839	
GDP			-0,122					0,755

Table 7: VARIMAX rotation.

VARIMAX rotation is similar to the results obtained from factor matrix, but the results of variables are different: in fact, the first component is related to education levels and children out of school. While the second one concerns long unemployment rate. The third is related to public percentage of GDP and the fourth, fifth and seventh refers to the importance of education in pre-primary, primary and secondary schooling. Finally, we calculate the factor transformation matrix. Below there are the results:

Factor	1	2	3	4	5	6	7	8
1	0,299	0,906	0,143	-0,217	-0,014	-0,123	0,007	0,084
2	0,500	-0,303	0,764	-0,067	-0,151	-0,098	0,105	-0,163
3	-0,637	0,239	0,573	0,272	0,189	0,266	-0,010	-0,168
4	0,241	-0,073	-0,034	-0,305	0,809	0,409	0,124	-0,078
5	0,120	0,049	-0,033	0,511	0,086	-0,021	0,770	0,346
6	0,407	0,068	-0,038	0,664	0,079	0,286	-0,543	0,072
7	0,021	0,002	0,020	-0,242	-0,480	0,783	0,101	0,295
8	0,129	0,133	-0,252	0,146	-0,209	0,204	0,274	-0,849

Table 8: factor transformation matrix.

In the factor transformation matrix we report results of factor matrix with VARIMAX rotation considering every principal component; thus, we show a little changing in the distribution of the phenomenon under analysis. Indeed, the influence of each variable is conditioned by presence of a large number of components which can explain the phenomenon itself. Multicollinearity of the variables is decreased by the application of factor analysis, but we note there is a dispersion of the phenomenon caused by some variables in the distribution. Once we calculated the components with maximum likelihood method we apply multivariate regression analysis using components resulting from factor analysis. In the next section we report results of this new analysis.

6. Multivariate regression model for literacy rate with maximum likelihood components

Factor analysis is helpful to use resulting components for regression multivariate analysis of literacy rate. The first analysis is conducted considering eight components. Following we show the results:

R	R square	adjusted R square	std. error of the estimate
0,981	0,963	0,950	0,4876017

Table 9: determination index.

constant	98,255 (0,085)	REGR factor score 5	0,429 (0,098)
REGR factor score 1	-0,053 (0,093)	REGR factor score 6	2,455 (0,104)
REGR factor score 2	-0,035 (0,089)	REGR factor score 7	-0,146 (0,102)
REGR factor score 3	0,041 (0,092)	REGR factor score 8	0,214 (0,112)
REGR factor score 4	-0,100 (0,101)		

Table 10: multivariate regression model for literacy rate with maximum likelihood components.

In table 9, we note a high value for determination index, which reaches 0,981. In previous analysis it reached only 0,223, hence this model is better than previous one. We consider also that in this model we include all components extracted by maximum likelihood method. We note that in table 10 we show a significant parameter only for component 5, 6 and 8 (besides the constant is also significant). The three components have a significance for a margin of error of less than 0,001; this result is satisfactory for our analysis. Consequently, we repeat multivariate regression analysis considering only these three components:

R	R square	adjusted R square	std. error of the estimate
0,978	0,957	0,953	0,4753965

Table 11: determination index.

Constant	8,255 (0,083)
REGR factor score 5	0,425 (0,096)
REGR factor score 6	2,461 (0,102)
REGR factor score 8	0,190 (0,108)

Table 12: multivariate regression model for literacy rate with maximum likelihood principal components

In table 11, we report the value of determination index of this analysis. The result is very similar to that reported in previous analysis, standing on 0,978 but it is related to the use of only three components. This result is the best analysis we made for the phenomenon of literacy rate respect to all variables and all components considered previously. Parameters are significant for all three components calculated with tolerance of less than 0,001. Overall, we report an excellent result of analysis, while these three components are sufficient to explain the phenomenon of literacy rate. Consequently we consider that component 5 regards education level, while component 6 is social exclusion of various countries, and 8 is economic component represented by GDP real growth rate. Below there is a chart which reassume values of literacy rate, expected and residual values of the last model:

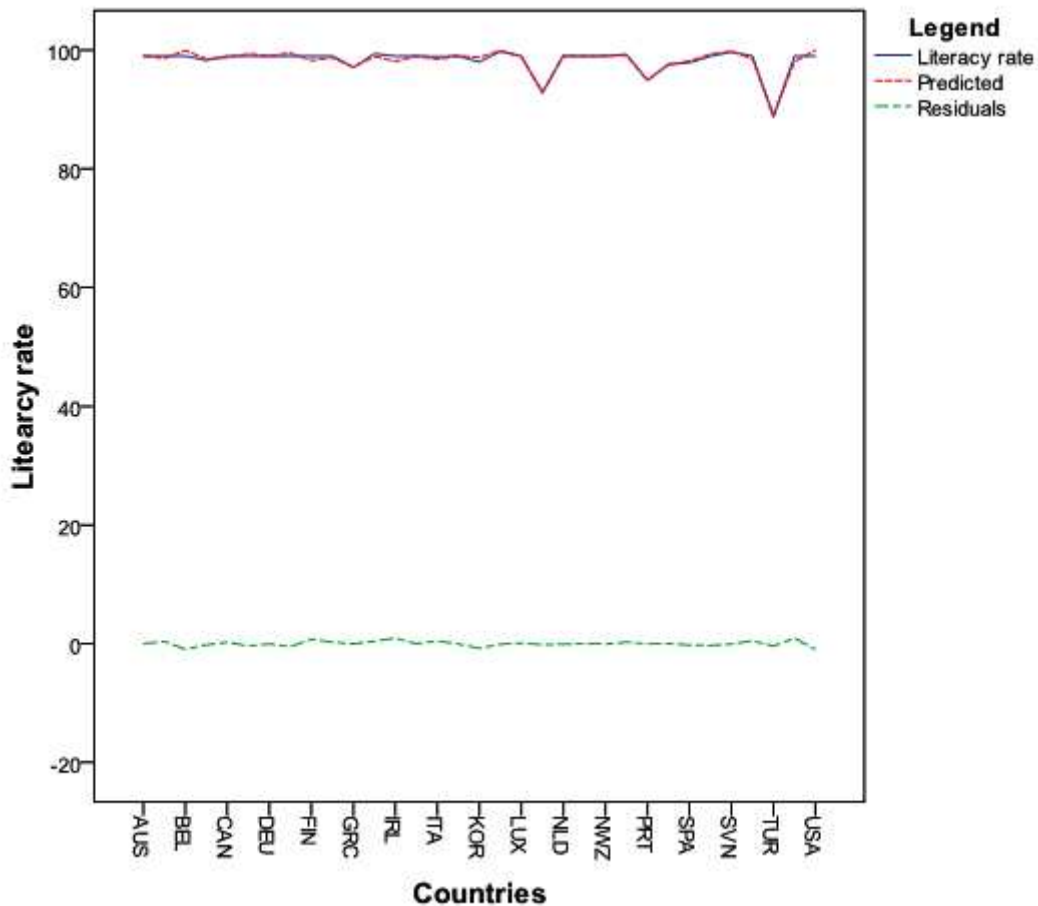


Figure 1: literacy rate model with observed, predicted and residual values.

In figure 1 it is clear that expected values of literacy rate are very similar than observed data. Determination index confirms an optimal approximation of theoretical to real data. Moreover, success of the model is confirmed by residual values which tends to zero and have very small variations, furthermore we note that residuals values are also distributed randomly: this indicate a good result of analysis. Finally on horizontal axis we report names of the countries: each value belongs to literacy rate values. In figure 1 we note three countries having a literacy rate below than other countries considered, and we can be see three light spikes downwards due to smaller values. The three countries are: Mexico, Portugal and Turkey. Having identified the presence of a lower literacy rate for these countries we have divided countries by a clustering to evidence differences between them.

7. Hierarchical cluster with maximum likelihood components for literacy rate

In this final section we analyze literacy rate with component 5 and 6 extracted in factor analysis with maximum likelihood method. For analysis we consider a hierarchical cluster, applying an Euclidean distance and a complete linkage clustering which are presented in methodology section as (12) and (13). By means of the dendrogram (not reported here) we find 3 important clustering. Thus, once divided countries in question in three clusters we report results graphically:

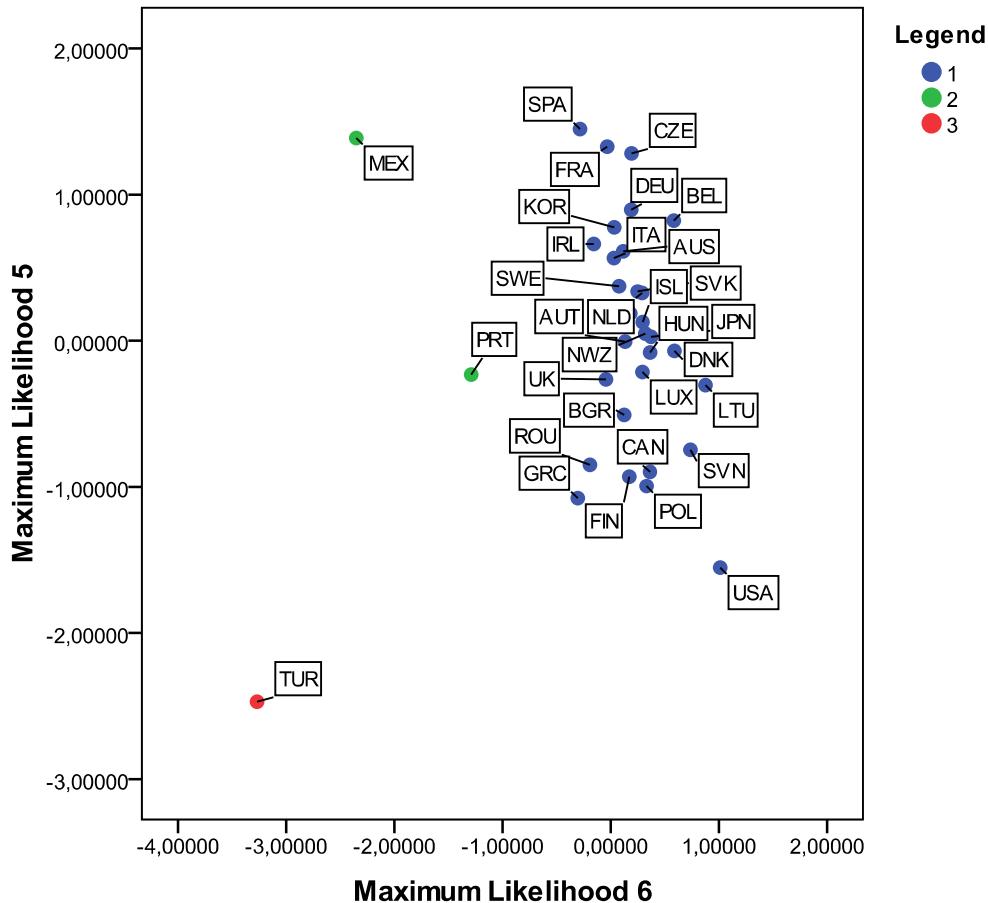


Figure 2: hierarchical cluster with maximum likelihood components for literacy rate.

Figure 2 shows that cluster analysis is applied to components 5 and 6 extracted by maximum likelihood method. The sixth component regards social exclusion variables such as long unemployment rate and children out of school, while the fifth component represents education levels which consists in pre-primary, primary, secondary and tertiary schooling. Figure 2 shows also a cluster consisting of a single country that is Turkey. This country presents the lowest literacy rate compared to other countries, and indeed in our figure is placed at the bottom left contemplating a lower level of education levels than other countries. A second cluster is composed by Mexico and Portugal, in which these countries are located in two different areas, because Portugal is closer to other countries, while Mexico is ranked higher than Portugal. Moreover, these two countries are in the same cluster because they have a lower literacy rate than other countries considered, and a lower

education levels. The third cluster is represented by majority countries, which report the highest levels of education and social exclusion standing above the three countries discussed previously. In conclusion we can affirm that literacy rate, considered as a factor of social exclusion, can be affected by social exclusion variables. This conditioning cannot occur early in the analysis, because of the heterogeneity of data has led to different results. As a consequence, overall both the results of multivariate regression model and those related to maximum likelihood method and cluster analysis confirm an exclusion of Turkey, Mexico and Portugal due to their relatively lower education than other countries concerned. The condition of relative low education levels headed to the last three countries leads them on a multidimensional process of progressive social disruption.

7. Concluding remarks

The aim of our study has been to examine how certain variables which characterize the various processes of social exclusion have an impact on literacy rates in terms of education in a large number of countries. It is known that the complex phenomenon of social exclusion has a great influence on the economic performance of a country through a variety of channels. We focus on the average education level of each country concerned, which as in the case of social capital, is influenced by several variables that can be traced to social exclusion itself. When the socio-economic conditions are difficult for individuals, such as widespread poverty and unemployment, it is more difficult to undertake and maintain optimal paths of education and training. This occurs more in Western countries, where competition is high and are increasingly practice of lifelong learning, and where human capital has a crucial role in the dynamics of growth and development.

The lack of an adequate literacy, relative to the average level of each social community, increases the risks of acquiring a working system protected, but also the need to learn to live civilly with others. These rules of social behaviour which are usually shared in the family context, as well as in that of school and neighborhood, and when they fail they come to represent a drive toward exclusion from the society which one belongs to. Literacy and education in particular are therefore a foundation of a civil society who wants to grow in terms of social, cultural, and economic prosperity and should be a national policy objective of promoting them, often through direct incentives to the weaker groups of society. The variables which we consider in order to observe the effects on the literacy rate, for the cross-country analysis we provide, cover the economic aspects of the school enrollment rates at various levels, but also public investment dedicated to education. We know that conditions which hinder educational processes may result from economic households difficulties, which in turn are often influenced by economic conditions, especially for people already at risk. It is also interesting to take into account the number of students enrolled at various educational levels as an indicator of widespread change in terms of behaviour toward education, as well as the economic outlook of households. According to the findings of an initial analysis, we do not verify a strong relationship of literacy with social exclusion: the basic problem is due to the heterogeneity of the data in the countries concerned. Another problem we face is the redundancy of information in the variables analyzed: we find the presence of multicollinearity which led us to apply a point estimate in the dataset.

However, factor analysis with maximum likelihood method has been useful to identify the main components in the estimate of literacy rate. It has been necessary to add a variance maximum rotation to transform parts of the model. Finally, we identify three components which can explain the overall dispersion of the phenomenon of literacy rate: then, we have obtained a phenomenon composed from nine initial variables to three unique variables. Finally we group the 30 countries in three key clusters which have naturally excluded those countries with education and social

differences than others. Countries that suffer exclusion are Mexico, the Turkey and Portugal. In these countries we note the presence of a literacy rate lower than any other and a very important element of social exclusion. Ultimately, our analysis has served to recognize the differences between countries both in terms of literacy at the social dimension. Indeed, we believe that social exclusion is one of many determinants of a country's literacy rate and evidence we study has confirmed empirical differences and heterogeneity of different countries, but we have identified also the presence of different continental reality as Australia, U.S. and some European countries which regarding literacy aspects have strong similarities.

References

Besset D. (2001), "Object-Oriented Implementation of Numerical Methods. An Introduction with Java & Smalltalk", Morgan Kaufmann.

Bhalla A., Lapeyre F. (1997). "Social Exclusion: Towards an Analytical and Operational Framework." *Development and Change*, Volume 28, Number 3, July 1997 , pp. 413-433(21)

de Gregorio J. and Lee J.-W. (2002), "Education and Income Inequality: New Evidence From Cross-Country Data", *Review of Income and Wealth*, Vol. 48, N. 3, pp. 395-416 (22).

Denny K. (2002). "New Methods for Comparing Literacy across Populations: Insights from the Measurement of Poverty." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 165, No. 3 (2002), pp. 481-493.

Freund R. (2003). "Statistical methods", Academic Press.

Galor O., Moav O. (2004). "From physical to human capital accumulation: inequality and the process of development." *Review of Economic Studies*, (2004) 71, 1001-1026.

Gradstein M., Justman M. (2002). "Education, Social Cohesion, and Economic Growth." *The American Economic Review*, Vol. 92, No. 4 (Sep., 2002), pp. 1192-1204.

Grossman H. I., Kim M. (1997). "Human Capital and Predation: A Positive Theory of Educational Policy." Working Papers 97-30, Brown University, Department of Economics.

Grossmann V. (2008). "Risky human capital investment, income distribution, and macroeconomics dynamics." *Journal of Macroeconomics*, 30 (2008) 19-42.

Hoover K. D. and Perez S. J. (2004), "Truth and Robustness in Cross-Country Growth Regressions", *Oxford Bulletin of Economics and Statistics*, Vol. 66, N. 5, p. 765.

Jakeman A. (2008), "Environmental Modelling, Software and Decision Support", Elsevier.

Kurland J. (1974). "Public Health in the New Millennium II: Social Exclusion." *Public Health Reports (1974-)*, Vol. 115, No. 4 (Jul. - Aug., 2000), pp. 298-301.

Lareau A., McNamara Horvat E. (1999). "Moments of Social Inclusion and Exclusion Race, Class, and Cultural Capital in Family-School Relationships." *Sociology of Education*, Vol. 72, No. 1 (Jan., 1999), pp. 37-53.

Levine R. and Renelt D. (1991), "Cross-Country Studies of Growth and Policy. Methodological, Conceptual, and Statistical Problems." Working Paper 608. Macroeconomic Adjustment and Growth Division, Country Economics Department, World Bank, Washington, D.C..

Levine R. and Renelt D. (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions", *American Economic Review*, Vol. 82, N. 4 (September).

Mardia K. and Kent J. (1980), "Multivariate analysis", Academic Press.

Moser W. (1996), "Linear models. A mean model approach", Academic Press.

OECD / Statistics Canada (2000), "Literacy in the Information Age: Final Report of the International Adult Literacy Survey", Paris: OECD.

OECD (2010), "Adult Literacy", Centre for Effective Learning Environments, Paris: OECD.

Oxenham J. (1980), "Literacy: Writing, reading, and social organization, *International Library of Welfare and Philosophy*", Routledge & Kegan Paul.

Porter F. (2000). "Social Exclusion: What's in a Name?" *Development in Practice*, Vol. 10, N. 1 (Feb., 2000), pp. 76-81.

Temple J., Johnson, P. A. (1998). "Social Capability and Economic Growth." *Quarterly Journal of Economics*, August 1998, 113(3), pp. 965-90.

UNESCO (1978), "Estimates and projections of illiteracy", *Current Study and Research in Statistics*, p. 154, Paris: UNESCO.