JRC Scientific and Technical Reports

ELLI-Index: a sound measure for lifelong learning in the EU

Michaela Saisana

EUR 24529 EN - 2010

JRC
EUROPEAN COMMISSION

Institute for the Protection and Security of the Citizen

The mission of the JRC-IPSC is to provide research results and to support EU policy-makers in their effort towards global security and towards protection of European citizens from accidents, deliberate attacks, fraud and illegal actions against EU policies.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: Michaela Saisana, European Commission, JRC, TP361, 21027, Italy.
E-mail: michaela.saisana@jrc.ec.europa.eu
Tel.: +39-0332-786572
Fax: +39-0332-785733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

Composite Indicators website: http://composite-indicators.jrc.ec.europa.eu/

Legal Notice
Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/

JRC 60268

EUR 24529 EN
ISBN 978-92-79-15629-8
ISSN 1018-5593
doi:10.2788/145

Luxembourg: Office for Official Publications of the European Union

© European Union, 2010

Reproduction is authorised provided the source is acknowledged

Printed in Italy

ELLI-Index: a sound measure for lifelong learning in the EU

Michaela Saisana

Executive Summary

The European Lifelong Learning Indicators (ELLI) project is an initiative led by the Bertelsmann Foundation, and one of its aims is to develop, test and pilot a new aggregate measure, the ELLI-Index, for country-level assessment of lifelong learning in the EU Member States. The ELLI-Index is developed with a view to be useful and accessible to a wide audience, including policy-makers, education researchers and practitioners, individual students and parents. Since its conception in late 2007, the ELLI has gone through a series of revisions and modifications based on the feedback received from Workshops and on iterative use of statistical tools, such as the ones presented in the present report.

The conceptual framework for the ELLI-Index (Bertelsmann Stiftung, 2010) is loosely based on the UNESCO's International Commission on Education for the Twenty-first Century (Delors et al, 1996) and the four major dimensions of learning identified: (a) Learning to Know (includes acquisition of knowledge and mastery of learning tools such as concentration, memory and analysis), (b) Learning to Do (concerns occupational, hands-on and practical skills), (c) Learning to Live Together (concerns learning that strengthens cooperation and social cohesion), and (d) Learning to Be (includes the fulfilment of a person, as an individual/member of a family/citizen). These four dimensions are not simply a conceptual framework, but a broader approach to measuring prosperity in a knowledge-based society.

The ELLI-Index combines 36 variables of lifelong learning, most of them coming from Eurostat. These variables reflect a wide range of learning activities, including participation rates in formal education and training, literacy skills (PISA), employees participating in CVT courses, labour market policies expenditure, and community engagement through cultural activities, among others. Besides seeking the appropriate sources of data and variables to build the four-dimensional framework, great emphasis was given in identifying the known economic and social outcomes of learning, such as income, employability, population health and social cohesion and democracy. These outcomes were perceived as components of the well-being of a society and were used as part of the statistical model that determined the relationship between the learning

inputs and the socioeconomic outcomes. The ELLI model employs factor regression analysis, with a view to ensure a statistically significant and high degree of correlation between the learning inputs and the learning outcomes ($r_S = 0.94$). However, the JRC analysis indicates that a similar association ($r_S = 0.92$) would have been obtained if one had used the classical equal weighting approach within and across the four learning dimensions.

The analysis and the subsequent recommendations of the present report follow the guidelines offered in the OECD (2008) Handbook on Composite Indicators and elicit from the lessons learnt from similar assessments carried out on other known composite indicators, such as the Environmental Performance Index, the Multi-dimensional Poverty Assessment Tool and the Index of African Governance.

The aims of this JRC validation report are: (a) to suggest eventual conceptual and methodological modifications in the ELLI-Index, (b) to assess the coherence of the ELLI conceptual framework, and (c) to test alternative models to build the Index.

The analysis of statistical coherence of the ELLI-Index is carried out at by applying multivariate statistical techniques to the underlying measures/variables within each of the four learning dimensions. Uncertainty and sensitivity analysis are next performed to evaluate the impact on the results of different models in which different sources of uncertainty are activated simultaneously. These scenarios differ from one another in the four-dimensional structure, the normalisation method of the raw measures/variables, the weights, and the aggregation formula within and across the four learning dimensions. This type of multi-modelling approach and the presentation of the ELLI-Index results under uncertainty, rather than as single numbers to be taken at face value, helps to avert the criticism frequently raised against composite measures, namely that they are generally presented as if they had been calculated under conditions of certainty, while this is rarely the case.

The overall assessment of the ELLI-Index reveals no particular shortcomings in the conceptual structure. In brief, the analyses demonstrate that the ELLI framework:

• is coherent from a conceptual and statistical point of view,
• has a well-balanced structure (not dominated by a single learning dimension), and
• is robust with respect to alternative normalisation, weighting and aggregation approaches.

In brief, this JRC report shows that the ELLI-Index is built according to a sound statistical methodology, its dimensions are well balanced and country ranking's dependence upon input assumptions does not exhibit any of the pathologies which at time affect composite measures.

However, recommendations for fine-tuning some data quality and methodological issues are made and summarised in the following:

Data quality issues:

- Two values in two variables – "GDP per capita", and "Anyone to discuss intimate and personal matters with" – need to be treated prior to applying a linear aggregation in the ELLI model.
- A note on "poor data coverage" needs to be added on two variables: "Involved in work for voluntary or charitable organization" in the Learning to Live Together dimension, and "Satisfaction with the job" in the socioeconomic outcomes.

Structural and modeling issues:

- A better measure of Environmental consciousness/awareness is needed, since the current variable on EPI environment is almost non-significantly correlated with the overall ELLI-Index.
- Two variables could eventually be assigned to different learning dimensions: (a) the "work-life balance" variable suits, statistically, better within the Learning to Live Together dimension (as opposed to the Learning to Be currently), (b) the "Labour market expenditure in training" is more correlated to the Learning to Live Together dimension as opposed to the Learning to Do as conceptualised. Hence, labour market spending in training appears to be more related to learning for social cohesion than to vocational learning.
- Eventually simplify the ELLI model by using equal weights within and across the four dimensions, since the results obtained with the two approaches are equivalent. The equal weighting approach has the further advantage that it is easier to communicate to a wide audience.

Dissemination of results:

- Four countries appear to be slightly misplaced in the overall ELLI-Index ranking – Estonia, Spain, Latvia and Slovakia. Any message drawn on the basis of the ELLI-Index for those four countries should be formulated with some caution due to the methodological assumptions made in developing the Index.

Hence, upon some refinements, the ELLI-Index can reliably be used to identify weaknesses and possible remedial actions, prioritize countries with relatively lower levels of lifelong learning conditions, and ultimately monitor and evaluate policy effectiveness. The ELLI-Index allows for the setting of national benchmarks in lifelong learning, and for further international comparisons of the underlying measures/variables of learning. At the same time, it allows for comparisons with other measures, such as competitiveness or innovation. In addition, the ELLI-Index highlights the link between learning and social cohesion and social cohesion and democracy. While many in education, health and other fields may have accepted the new paradigms of lifelong learning, the ELLI-Index is likely to open up these ideas and dialogues to a wider population within and outside Europe.

Table of Contents

List of Tables

List of Figures

1. Introduction

The European Lifelong Learning Indicators (ELLI) project is an initiative led by the Bertelsmann Foundation, and one of its aims is to develop, test and pilot a new aggregate measure, the ELLI-Index, for country-level assessment of lifelong learning in the EU Member States. The conceptual framework for ELLI (Bertelsmann Stiftung, 2010) is loosely based on the UNESCO's International Commission on Education for the Twenty-first Century (Delors *et al.*, 1996) and the four major dimensions of learning identified: (a) Learning to Know (includes acquisition of knowledge and mastery of learning tools such as concentration, memory and analysis), (b) Learning to Do (concerns occupational, hands-on and practical skills), (c) Learning to Live Together (concerns learning that strengthens cooperation and social cohesion), (d) Learning to Be (includes the fulfilment of a person, as an individual/member of a family/citizen). These four dimensions are not simply a conceptual framework, but a broader approach to measuring prosperity in a knowledge-based society. The ELLI-Index combines 36 measures/variables of lifelong learning coming mainly from Eurostat. These measures/variables reflect a wide range of learning activities, including participation rates in formal education and training, literacy skills, employees participating in CVT courses, labour market policies expenditure, and community engagement through cultural activities, among others.

The four-dimensional framework is accompanied by selected indicators on known economic and social benefits of learning, such as income, employability, population health and social cohesion and democracy. These outcomes were perceived as components of the well-being of a society and were used as part of the statistical model underlying the ELLI-Index. The ELLI model employs factor regression analysis in order to estimate the weights to be assigned to the measures/variables within each dimension and to the four dimensions, so as to maximise the degree of correlation between, on one side, each of the four dimensions or their overall aggregate, the ELLI-Index, and the socioeconomic outcomes of learning on the other. Prior to aggregation, the raw data were standardised (subtracting the variables mean and dividing by the standard deviation). This standardisation was applied at both levels of aggregation from the variables to the four dimensions, and from the four dimensions to the overall ELLI-Index.

The present study aims to critically assess the methodological approach taken by the Bertelsmann Foundation to build the ELLI-Index, by addressing two key questions:

- Is the ELLI framework both statistically and conceptually coherent?

- What other models could be used to build the ELLI-Index?

The analysis and the subsequent recommendations of the present report follow the guidelines offered in the OECD (2008) Handbook on Composite Indicators and elicit from the lessons learnt from similar assessments carried out on other known composite indicators, such as the Environmental Performance Index[1], the Multi-dimensional Poverty Assessment Tool[2], the Index of African Governance[3], and the Composite Learning Index[4].

The report is structured as follows. **Section 2** describes the ELLI-Index conceptual framework (dimensions, indicators, measures and socioeconomic outcomes), and the methodological approach used to build the ELLI-Index. **Section 3** discusses data quality issues (missing data, eventual outliers) and suggests some fine-tuning. **Section 4** deals with eventual refinements in the Conceptual Framework based on an analysis of the correlation structures within and across dimensions. In **Section 5**, we carry out an uncertainty and sensitivity analysis of the ELLI-Index. We aim to examine to what extent the country ranking of the EU Member States depends on the choices made during the development of the ELLI-index. The analysis involves the simultaneous activation of various sources of uncertainty (e.g. preserving or not the four-pillar structure, normalisation of raw data, weights, and aggregation formula). **Section 6** concludes.

---

[1] Saisana M., and Saltelli A., 2010, Uncertainty and Sensitivity Analysis of the 2010 Environmental Performance Index, EUR 56990 EN, European Commission- JRC-IPSC, Italy.

[2] Saisana M., and Saltelli A., 2010, The Multidimensional Poverty Assessment Tool (MPAT): Robustness issues and Critical assessment, EUR 24310 EN, European Commission- JRC-IPSC, Italy.

[3] Saisana M., Annoni, P, Nardo M., 2009, A robust model to measure African Governance: Robustness Issues and Critical Assessment, EUR 23274 EN, European Commission, JRC-IPSC, Italy.

[4] Saisana M., 2008, The 2007 Composite Learning Index: Robustness Issues and Critical Assessment, EUR 23274 EN, European Commission, JRC-IPSC, Italy.

2. Conceptual framework

In most European countries, lifelong learning is being promoted in response to forces of globalization and the imperative to create a "Knowledge Society" for all (European Civil Society, 2004). Important questions in making public policy for lifelong learning include: How much lifelong learning is going on? Of what type and where? Is this the right amount and type of lifelong learning? Do we need more, or less, of certain types of lifelong learning? Answering these questions presumes the ability to measure lifelong learning.

While there is a plethora of indicators that describe various aspects of learning, none of them individually suffices to measure the intangible concept of lifelong learning. However, a single summary measure of lifelong learning could make it possible to assess whether things are getting better or worse; it would allow the general public and media to follow and monitor one number rather than tens of indicators; it would help contribute to priority setting and policy formulation; and lastly, it would also make it easier to compare trends over time and across countries. Thus, an index of lifelong learning in the EU countries could reveal new knowledge which otherwise would remain invisible.

Thus far, no comprehensive measure of lifelong learning in Europe exists. This may be due to the fact that learning represents a minefield of conceptual and methodological questions, and is thus difficult to define, isolate, measure and apply empirically (Levy, 1994). In other disciplines, such as economy or environment, aggregate measures (or composite indicators) of performance are popular tools for presenting complex concepts (Bandura, 2008). Nevertheless, many authors warn against the dangers of the misuse of composite measures, as the numbers are often taken at face value with little discussion of their validity. Successful composite indicator analysis depends on mastering the art of indicator selection and composite indicator design. For reasons of public accountability, composite indicators – as with any advanced evaluation method – must have a clear and transparent structure and be based on accepted concepts.

Canada, however, has pioneered such a composite indicator that attempts to describe lifelong learning across the more than 4500 communities in the country from 2006 till 2010 (Canadian Council on Learning, 2010). The Composite Learning Index by the Canadian Council on Learning has been an inspiration to the ELLI-Index by the Bertelsmann Foundation and its international team of experts.

Four dimensions of learning

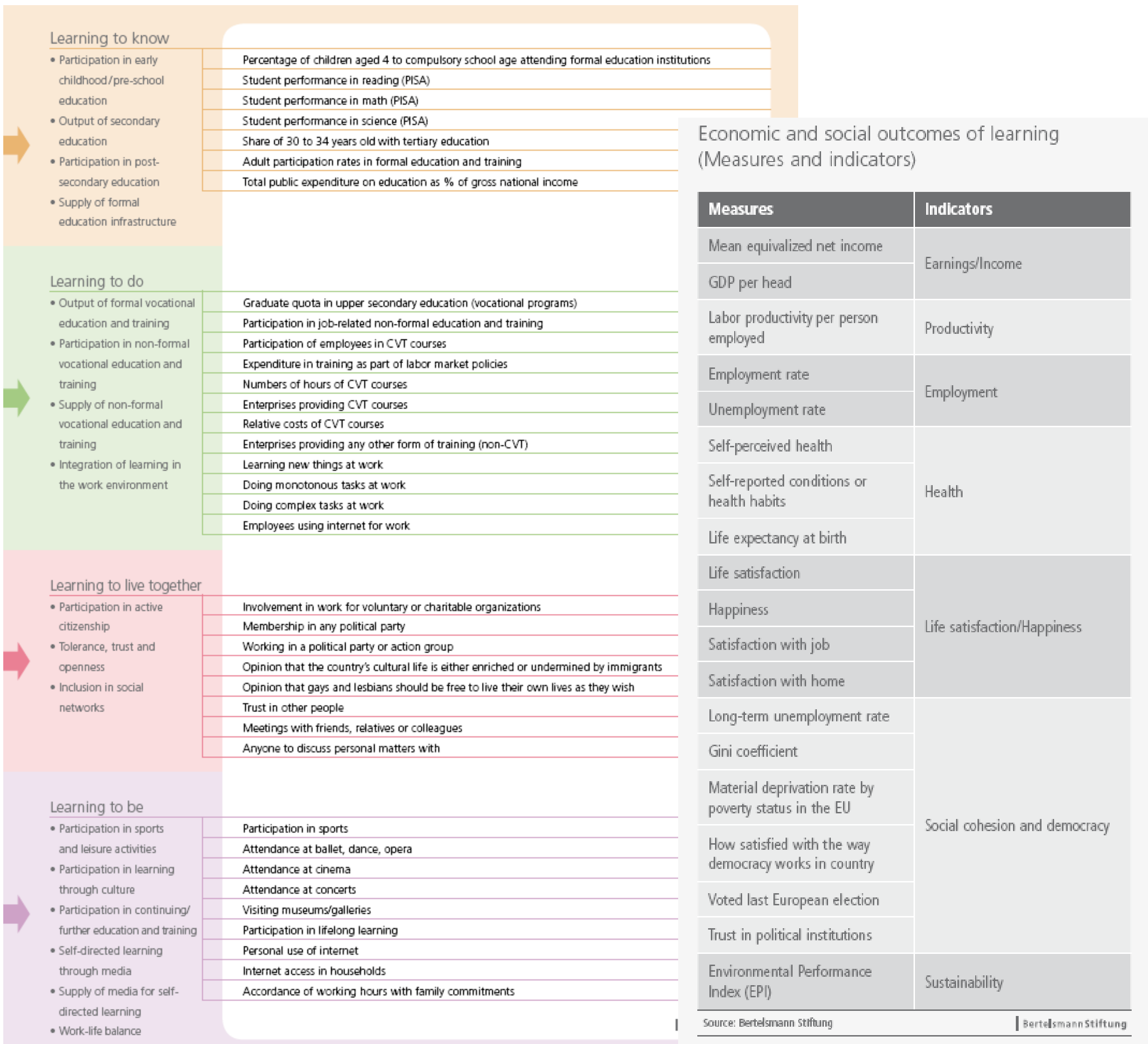
The conceptual framework of the ELLI-Index (Figure 1), is loosely based on the recommendations by the UNESCO International Commission on Education for the Twenty-first Century on the four-dimensions of learning: learning to know (essentially school-based learning), learning to do (learning related to work and vocational skills), learning to be (learning that relates to personal development and creativity), and learning to live together (learning that relates to social cohesion and participation in communities) (Delors *et al.*, 1996). In the intervening decade they have become a crucial point of reference in education worldwide. The life-wide and lifelong perspectives on the learning concept are encapsulated in this framework in an imaginary two-dimensional plane of time versus space on which to consider learning contexts beyond formal education (Dave, 1976; Aspin *et al.*, 2001). On the time axis, and hence relating to the term "lifelong", there are the successive stages in which learning occurs, such as early childhood, primary, secondary and tertiary schooling, as well as adulthood. On the space axis, and hence expressing "life-wide", there are the different contexts at every stage of life, such as at home, school, work, community, leisure, etc.

In line with these four learning dimensions, and on the basis of expert opinion, literature review and correlation analysis, 36 measures/variables were selected as most relevant in describing lifelong learning in the European Union Member States (Bertelsmann Stiftung, 2010). The Learning to Know dimension includes indicators on early childhood education, outputs of secondary education (PISA), participation in post-secondary education and formal education infrastructure. The Learning to Do dimension is composed of outputs of formal vocational education and training (VET), participation in continuous VET, supply of continuous VET and integration of learning in the work environment. The Learning to Live Together dimension covers participation in active citizenship, tolerance and openness, trust in people and inclusion in informal social networks. Finally, the Learning to Be dimension includes indicators on participation in sporting and leisure activities, in cultural life/learning, in continuing further education/training, self-directed learning through media and work-life balance.

Socioeconomic outcomes of learning

The four dimensions of learning and the selected learning inputs can only partly describe lifelong learning in the European Union Member States. Another aspect of the concept is captured by the socioeconomic outcomes of learning (Figure 1). Fourteen social outcomes of learning (related to health, life satisfaction, social cohesion and democracy, and sustainability) and five economic outcomes of learning (related to earning/income, productivity, employment) were selected in order to determine the strength of the relationship between learning inputs and socioeconomic outcomes.

Figure 1. Four-dimensional framework for lifelong learning in Europe and socioeconomic outcomes of learning



Source: Bertelsmann Stiftung (2010) Making Lifelong Learning Tangible – The European ELLI-Index 2010.

The ELLI-Index model

The development of the ELLI-Index was based on the broad concept of measurement validity in the field of educational measurement (Brennan, 2006). The methodology also requires the explicit specification of those outcomes to which a society aspires which are also conceptually dependent on the indicators and measures underlying the lifelong learning model. Thus, the statistical construct representing the social and economic outcomes is the external criterion against which the ELLI-Index model was optimized.

Among the various regression-based methods, factor analysis regression was chosen for the development of the ELLI-Index. Factor analysis regression allows for greater conceptual control than principal components regression or partial least squares regression, by providing meaningful interim estimates of the latent variables – in this case, the four pillars of learning (Jolliffe, 1982; Helland, 1992). Furthermore, factor analysis regression is less restrictive than structural equation modelling (SEM), which requires all model parameters to be estimated simultaneously against a common covariance matrix (Kaplan, 2000). The successive stages of factor analysis regression allows for the development of models which with SEM would be unidentifiable or impractical. The model underlying the ELLI-Index is summarised in six steps:

Step 1. The country scores for the 36 measures/variables learning (Figure 1) were first adjusted so that in all cases higher values corresponded to higher levels of lifelong learning (e.g. doing monotonous tasks at work) and then standardised ($mean = 50$, $std = 10$) to equalize differences in the variation of the different measures.

Step 2. Factor analysis (FA) was applied to the variables within each dimension to extract those orthogonal (uncorrelated) factors that cumulatively explained at least 90% of the variance within the dimension (see Section 4 for more details).

Step 3. FA was also applied to extract a single common factor from the 19 socioeconomic outcomes of learning (see Section 4 for more details).

Step 4. Ordinary least squares regression (OLS) was employed to estimate the weights of the factors within each dimension, so that each dimension scores would have the highest association with the common factor of the socioeconomic outcomes.

Step 5. The country scores on the four dimensions were next standardized and principal components analysis was used to transform them into four orthogonal (uncorrelated) dimensions.

Step 6. Finally, OLS was employed to estimate the weights of the four dimensions, so that their weighted arithmetic average, the ELLI-Index, would have the highest association with the common factor of the socioeconomic outcomes.

Missing data were replaced with the most recent year available (up to three years). Remaining missing data were estimated using the Expectation-Maximization (EM) algorithm. The EM algorithm (Dempster, Laird, and Rubin, 1977; Little and Rubin, 1992) is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating the following steps:

1. The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates.

2. The maximization M-step: Given a complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step.

The two steps are iterated until the iterations converge.

The model used to build the ELLI-Index relied on statistical analysis and aimed to bypass decisions on the weighting issue in particular. An equal weighting scheme of the measures/variables or the dimensions was not selected by the developers because, although there is a strong basis for the theoretical involvement of each indicator in lifelong learning, there is no reason to suppose that their roles are equal. Given the availability of social and economic outcomes of learning, it was natural that regression weighting could be used.

Since composite indicator development is an art, it would not be prudent to argue *ex ante* that the model underlying the ELLI-Index is the best approach to measure lifelong learning at national level in Europe. For this reason, we will try to anticipate criticism by employing a multi-modelling approach, described in Section 5.

3. Data quality issues

3.1. Reproducing the ELLI-Index results

Transparency to stakeholders is considered to be an essential ingredient of well-built composite indicators (OECD, 2008). A clear understanding of the ELLI-Index methodology is also necessary with a view to perform the robustness assessment of the index. Thus, the first test was to try to reproduce the ELLI-Index results given the data and information provided to the public. We succeeded in doing so, as the relevant documentation provided in the relevant website www.elli.org provides enough information to a statistically literate public in order to replicate the methodology and the results. The ELLI-Index is clear about its definition, its framework, its underlying indicators and measures, its methodological assumptions, and does not fall under the critiques of normative ambiguity at times addressed to composite indicators (see Stiglitz report, p. 65).

3.2. Asymmetric distributions and outlier detection

We next assessed the appropriateness of using the standardization method to normalise the raw data. The standardization method (subtracting the variable mean and dividing by the standard deviation) is in general sensitive to outliers, which, if not treated properly, could become unintended benchmarks. Furthermore, outliers can have a strong impact on the correlation structure (see analysis in Section 4), and hence introduce bias in the ELLI model (which is based on correlations) and in the subsequent interpretation of the results. There are many methods suitable for outlier detection, but in the context of composite indicator building the combined use of skewness and kurtosis could be particularly apt. A skewness value greater than 1 together with a kurtosis value greater than 3.5 (both in absolute terms) could flag problematic indicators that need to be treated before the final index construction (Groeneveld and Meeden, 1984).

Only two variables are flagged for further consideration as they exhibit relative high values for skewness and kurtosis (Table 1): "Anyone to discuss intimate and personal matters with" in the Learning to Live Together dimension, and GDP per capita in the socioeconomic outcomes. Potential outliers could be identified either visually (as shown in Figure 2) or using information based on the inter-quartiles range, namely outside the range:

$$\begin{aligned} &\text{Lower boundary: } L = Q_1 - 1.5 \cdot (Q_3 - Q_1) \\ &\text{Upper boundary: } U = Q_3 + 1.5 \cdot (Q_3 - Q_1) \end{aligned} \qquad (1)$$

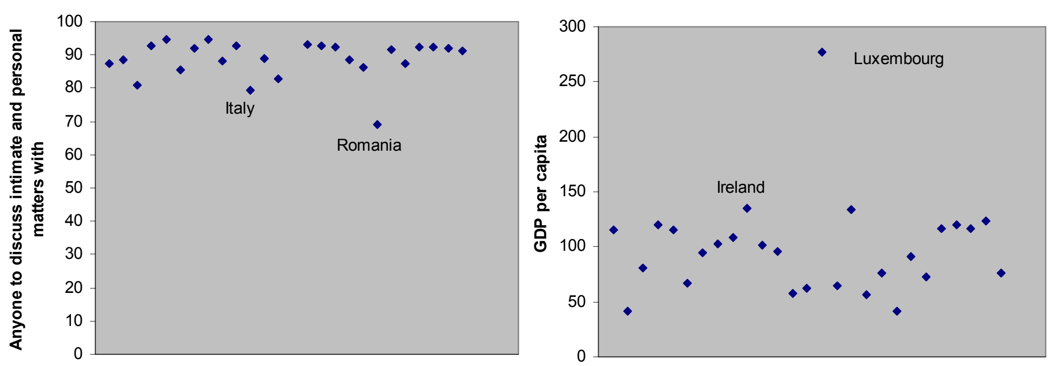
where Q_1 and Q_3 are respectively the first and the third quartile (Tukey, 1977). We will briefly refer to this method as the inter-quartiles range.

Both the visual approach and the inter-quartiles range spot the same outlier values: a single value for either indicator. For GDP per capita, the value 276.4 for Luxembourg is very high compared to the values for the remaining countries (Ireland's second best value is merely 135.4). Similarly, for "Anyone to discuss intimate and personal matters with", the value 69.2 for Romania is very low compared to the values for the remaining countries (Italy's second low value is 79.5). Given that only one value was identified as outlier in these two indicators, we have preferred not to apply any transformation (e.g., taking logarithms, Box-Cox, or other), but simply to winsorize the outlier values by resetting them to the second best/low value as shown in Table 1.

Table 1. Outlier detection and treatment

Variable/Measure	***N***	***Skewness***	***Kurtosis***	***Outlier treatment***
Anyone to discuss intimate and personal matters with (Learning to Live Together dimension)	25	-1.892	4.444	Value 69.2 for Romania set to 79.5 (= Italy)
GDP per capita (Outcomes)	27	2.314	8.838	Value 276.4 for Luxembourg set to 135.4 (= Ireland)

Figure 2.Problematic indicators (outliers)



Data quality tests focused next on availability at all levels: variables, dimensions, countries. The 2010 ELLI dataset is characterized by overall excellent data coverage (96%, matrix of (36+19) ×27, Table 2). The most complete dimension is the Learning to Be (only 0.8% missing values), the least complete is the Learning to Live Together (10.2% missing values), which is still acceptable according to some rules of thumb for data availability of at least 75-80%. The socioeconomic outcomes are also very well covered (only 2.7% missing values).

At the country level, data coverage is overall very good, but there are few countries with notable data gaps in some of the learning dimensions (Table 3). On the positive side, thirteen of the EU27 countries – Austria, Belgium, Estonia, Finland, Germany, Hungary, Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain and Sweden − do not miss any of the 36+19 values needed to build the ELLI-Index. There are, however, a few countries that miss more than 20% of the values in a given dimension. This is the case for Malta in the Learning to Know and Learning to Live Together dimensions, Lithuania in the Learning to Live Together dimension, and Cyprus and Ireland in the Learning to Know dimension. The decision of the ELLI development team not to present scores on these dimensions for those four countries is justified.

At the variable level, 19 measures of the 36 in the four learning dimensions and 15 measures of the 19 in the socioeconomic outcomes do not miss a single value (Table 4). However, two measures miss values for almost one-third of the countries: "Involved in work for voluntary or charitable organization" in the Learning to Live Together dimension, and "Satisfaction with the job" in the socioeconomic outcomes. According to general guidelines for composite indicator development, one should eliminate these two measures from the calculation of the Index. In the present case, given that the Index is made of 36+19 measures, eliminating these two measures would leave the results practically unaffected. It is recommended that the two measures are maintained in the conceptual framework but a note on poor data coverage is added.

Table 2. Missing data issues- dimension level

Learning dimension	*Missing data*	*Number of variables*	*Missing data (%)*
Learning to Know	11	7	5.8%
Learning to Do	8	12	2.5%
Learning to Live Together	22	8	10.2%
Learning to Be	2	9	0.8%
Total			
Outcomes	14	19	2.7%

Table 3. Missing data issues –country level

Countries with missing data	*Missing values in the four learning dimensions (total of 36 variables)*	*Missing values in the outcomes (total of 19 variables)*
Malta[LK,LTL]	14	3
Lithuania[LTL]	9	3
Ireland[LK]	4	0
Romania	3	2
Luxembourg	3	1
Cyprus[LK]	3	0
Czech Republic, Greece, Italy, Latvia	1	1
Denmark, France, United Kingdom	1	0
Bulgaria	0	1

Notes:
[LK] results on the Learning to Know dimension will not be presented for Malta, Cyprus and Ireland due to more than 20% missing data (rule of thumb used for the construction of ELLI)
[LTL] results on the Learning to Live Together dimension will not be presented for Malta and Lithuania due to more than 20% missing data.

Table 4. Missing data issues – variable level

Learning dimension	***Variables with missing data***	***Missing values (total of 27 countries)***
BE	Personal use of internet	2
DO	Participation in job related non-formal education and training	5
DO	Participation of employees in CVT courses	2
DO	Graduate quota in upper secondary education - pre-vocational and vocational programmes	1
KNOW	Adult participation in formal education and training	4
KNOW	Student performance in reading (PISA)	2
KNOW	Student performance in mathematics (PISA)	2
KNOW	Student performance in science (PISA)	2
KNOW	Percentage of children aged 4 to compulsory school age attending formal education institutions	1
LIVE	Involvement in work for voluntary or charitable organisations	8
LIVE	Trust in other people	2
LIVE	Meeting with friends, relatives or colleagues	2
LIVE	Anyone to discuss intimate and personal matters with	2
LIVE	Membership in any political party	2
LIVE	Working in political party or action group	2
LIVE	Opinion that the country's cultural life is either enriched or undermined by immigrants	2
LIVE	Opinion that gay men and lesbians should be free to live their own lives as they wish	2
Outcomes	Satisfaction with the job	8
Outcomes	Self-reported conditions or health habits	2
Outcomes	Satisfaction with the way democracy works in country	2
Outcomes	Trust in political institutions	2

4. Conceptual and statistical coherence of the ELLI framework

The "making of" the ELLI-Index demands a sensitive balance between simplifying lifelong learning aspects and still providing sufficient detail to detect characteristic differences between the EU countries. Such conflicting demands could finish by producing an aggregate measure that is almost impossible to verify, particularly since lifelong learning cannot be measured directly. It is therefore taken for granted that the ELLI-Index cannot be tested on the basis of ground truth.

Yet, in order to enable informed policy-making and to be useful as policy and analytical assessment tool, the Index needs to be assessed with regard to its validity and potential biases. The research question to be answered is:

- *Is the ELLI-Index coherent from a conceptual and statistical point of view?*

4.1 Statistical dimensionality of the four-dimensional framework

The major goal of this analysis is to let the data speak: that is, to assess whether the ELLI-Index conceptual framework is supported by the collected data. First, we study how many latent factors exist within each of the learning dimensions, and second, whether the four learning dimensions share a single or more latent factors.

By applying Principal Components Analysis (PCA) within a dimension and looking at the number of eigenvalues that are greater than roughly 1.0 according to the Kaiser criterion (assumption relaxed to greater than 0.9) (Manly, 1994; Dunteman, 1989) we notice that all dimensions can be summarized by at least two orthogonal latent factors (Table 5). The Learning to Be dimension is the most consistent dimension (highest degrees of correlation among its underlying variables) and the first factor captures roughly 64% of the variance in this dimension (the first factor in each of the other dimensions captures between 46% and 57%). The ELLI model was developed using all those factors within each dimension that account for at least 90% of the variance in each dimension, namely five orthogonal factors in Learning to Be dimension, six orthogonal factors in Learning to Do dimension, and four orthogonal factors in each of the Learning to Know and Live Together dimensions. These results imply that the selected variables in the ELLI-Index capture distinct and diverse aspects of lifelong learning, with considerable overlap of information.

The four learning dimensions can be summarized by a single latent factor that captures almost 85% of the variance of the four dimensions (Table 6). In fact, the Pearson correlation coefficients among the four learning dimensions are between 0.72 (Learning to Know with Learning to Do) and 0.86 (Learning to Be and Learning to Live Together). These results suggest that the four learning dimensions in the ELLI-Index share a lot of common information, which can not easily be isolated and studied independently. Furthermore, given that a single latent factor is identified among the four dimensions suggests that a linear aggregation rule can be used to aggregate the four dimensions, as was done in the ELLI model.

Table 5. Statistical dimensionality within the four learning pillars in ELLI

Principal Component	*Learning to Be*		*Learning to Do*	
	Eigenvalue	Cumulative variance (%)	Eigenvalue	Cumulative variance (%)
1	5.7	63.8	6.2	51.9
2	0.9	73.3	1.3	62.9
3	0.7	80.9	1.2	72.6
4	0.5	86.7	0.8	79.1
5	0.4	91.4	0.7	85.1
6	0.3	94.9	0.6	89.7
7	0.2	97.0	0.5	93.8
8	0.2	98.7	0.3	96.3
9	0.1	100.0	0.2	97.7
10			0.1	98.8
11			0.1	99.7
12			0.0	100.0
	Learning to Know		*Learning to Live Together*	
	Eigenvalue	Cumulative variance (%)	Eigenvalue	Cumulative variance (%)
1	4.0	56.6	3.7	46.2
2	1.0	70.7	2.0	70.6
3	0.9	83.7	0.9	81.3
4	0.7	93.2	0.7	89.6
5	0.4	98.5	0.4	94.1
6	0.1	99.3	0.3	97.3
7	0.0	100.0	0.2	99.3
8			0.1	100.0

Table 6. Statistical dimensionality of the four learning pillars in ELLI

Principal Component	*Eigenvalue*	*Variance (%)*	*Cumulative variance (%)*
1	3.39	84.84	84.84
2	0.30	7.47	92.31
3	0.20	5.04	97.35
4	0.11	2.65	100.00

4.2 Statistical dimensionality of the socioeconomic outcomes of learning

The first principal component of the 19 socioeconomic outcomes of learning has an eigenvalue of 10.6 and an explanatory power of 56% of the variance in the outcomes dataset (Table 7). The second important principal component accounts for an extra 9.3% of the remaining variance. Considering this sudden drop in the explanatory power, the developers decided to retain only the first factor as a summary measure of the socioeconomic outcomes.

The correlation between the aggregate measure of the socioeconomic outcomes and each of the underlying measures is shown in Table 8 in decreasing order of correlation (in absolute terms). Two social ("How satisfied with the way democracy works in country", and "Life satisfaction") and one economic measure ("Mean equivalised net income") capture more than 80% of the variance in the socioeconomic outcomes scores for the EU countries. High is also the explanatory power of most of the social measures, such as material deprivation by poverty status, trust in political institutions, happiness, satisfaction with the home, health. Also, the GDP per capita is strongly correlated with the socioeconomic outcomes. Almost non-significant is the correlation between the socioeconomic outcomes and the EPI environment. Note that the EPI environment is almost non-significantly correlated with the overall ELLI-Index. This result suggests that if the aim is to include a measure on environment that has an impact in the ELLI-Index, the EPI environment measure needs to be replaced by another measure that captures better environmental consciousness /awareness.

Table 7. Statistical dimensionality of the socioeconomic outcomes of learning

Principal Component	*Eigenvalue*	*Variance (%)*	*Cumulative variance (%)*
1	10.6	56.0	56.0
2	1.8	9.3	65.3
3	1.4	7.3	72.5
4	1.2	6.5	79.0
5	1.0	5.0	84.1
6	0.7	3.9	88.0
7	0.6	3.3	91.2
8	0.4	2.2	93.4
9	0.4	1.9	95.3
10	0.3	1.4	96.7
11	0.2	0.9	97.6
12	0.1	0.8	98.4
13	0.1	0.6	98.9
14	0.1	0.4	99.3

Principal Component	*Eigenvalue*	*Variance (%)*	*Cumulative variance (%)*
15	0.0	0.3	99.6
16	0.0	0.2	99.8
17	0.0	0.1	99.9
18	0.0	0.1	100.0
19	0.0	0.0	100.0

Table 8. Correlation between the socioeconomic measures and their aggregate

Type	*Measure in the socioeconomic outcomes*	*Pearson correlation coefficient with the aggregate socioeconomic outcomes*
Economic	Mean equivalised net income	.929
Social	Satisfaction with the way democracy works in country	.914
Social	Life satisfaction	.892
Economic	Labour productivity per person employed	.878
Social	Material deprivation rate by poverty status in the EU	-.877
Social	Trust in political institutions	.873
Social	Happiness	.859
Social	Satisfaction with the home	.858
Social	Self-reported conditions or health habits	.854
Social	Life expectancy at birth	.831
Economic	GDP per capita	.812
Social	Satisfaction with the job	.790
Social	Self-perceived health	.724
Social	Gini coefficient	-.539
Social	Voted last European election	.522
Economic	Employment rate	.503
Social	Long-term unemployment rate	-.478
Economic	Unemployment rate	-.414
Social	EPI Environment	.202

Note: Pearson correlation coefficients less than 0.4 are not statistically significant.

4.3. Cross-correlations between variables and dimensions

Next, we test whether the measures/variables are "statistically" assigned to the same dimension as conceptualised. A simple, but nevertheless informative approach to do so is by means of cross-correlation analysis between the variables and the four dimensions. Intuitively, one would expect that a variable is more correlated to its own dimension than to any of the other three dimensions. In most cases, the expectation that the variables are more correlated to their conceptual dimension than to any of the other three dimensions of learning is confirmed and furthermore all correlations are statistically significant and have the expected sign (Table 9).

There are two exceptions to this expectation worthy of further discussion. First, the measure on "work-life balance" is more correlated to the Learning to Do or Learning to Live Together dimension. Conceptually, the Learning to Live Together dimension captures learning for social cohesion, hence it may be suitable to move the measure on work-life balance from the Learning to Be to Learning to Live together dimension. Second, the "Labour market expenditure in training" is more correlated to the Learning to Live Together dimension as opposed to the Learning to Do as conceptualised. Hence, labour market spending in training appears to be more related to learning for social cohesion than to vocational learning. An eventual shift of this measure could be considered.

Furthermore, almost all 36 variables in the conceptual framework affect the ELLI-Index scores and hence one can confidently argue that "almost all what is included in the ELLI-Index has a saying on the results"[5]. There are only five variables (Table 9) that do not have a statistically significant linear association to the ELLI-Index scores (both Pearson and Spearman rank correlations are <0.4), they are however significantly, albeit low correlated with their own dimension, and hence the developers decided to keep them in the ELLI model.

Table 9. Correlations between indicators and dimensions, ELLI-Index and outcomes

Dimension	*Measure*	*Know*	*Do*	*Live*	*Be*	*ELLI-Index*
KNOW	Adult participation in formal education and training	0.72	0.59	0.49	0.55	0.62
KNOW	Student performance in reading (PISA)	0.78	0.64	0.52	0.68	0.69
KNOW	Student performance in mathematics (PISA)	0.77	0.73	0.57	0.72	0.74
KNOW	Student performance in science (PISA)	0.7	0.64	0.46	0.6	0.63
KNOW	Total public expenditure on education	0.82	0.6	0.54	0.66	0.69
KNOW	Percentage of children aged 4 to compulsory school age attending formal education institutions	0.37	0.35	0.39	0.47	0.43
KNOW	Share of 30-34 years old with tertiary attainment	0.86	0.58	0.7	0.79	0.78
DO	Participation in job related non-formal education and training	0.39	0.66	0.46	0.45	0.53
DO	Doing complex tasks at work	0.16	0.43	0.25	0.23	0.29
DO	Using internet at work	0.92	0.81	0.8	0.87	0.91
DO	Number of hours of CVT courses	0.59	0.82	0.66	0.73	0.75
DO	Graduate quota in upper secondary education - pre-vocational and vocational programmes	0.08	0.51	0.36	0.21	0.32
DO	Learning new things at work	0.78	0.83	0.72	0.8	0.84

[5] Note that this is not always the case. The inclusion of a variable in a conceptual framework provides no guarantee that the variable will affect the final Index results. This argument is an important remark to make as this is a common misconception among stakeholders that wish to have a saying on an Index by suggesting which variables to include.

Dimension	Measure	Know	Do	Live	Be	ELLI-Index
DO	Doing monotonous tasks	-0.29	-0.42	-0.38	-0.4	-0.4
DO	Participation employees in CVT courses	0.47	0.8	0.55	0.58	0.64
DO	Labour market expenditure in training	0.46	0.45	0.7	0.47	0.57
DO	Enterprises providing CVT courses	0.75	0.95	0.78	0.87	0.9
DO	Enterprises providing any other form of training	0.6	0.84	0.6	0.65	0.72
DO	Relative costs of CVT courses	0.58	0.75	0.49	0.68	0.66
LIVE	Trust in other people	0.85	0.81	0.83	0.88	0.9
LIVE	Involvement in work for voluntary or charitable organisations	0.51	0.62	0.77	0.62	0.69
LIVE	Meeting with friends, relatives or colleagues	0.65	0.61	0.86	0.72	0.78
LIVE	Anyone to discuss intimate and personal matters with	0.53	0.39	0.49	0.54	0.53
LIVE	Membership in any political party	0.1	0.39	0.51	0.28	0.36
LIVE	Working in political party or action group	0.07	0.36	0.51	0.24	0.34
LIVE	Opinion that the country's cultural life is either enriched or undermined by immigrants	0.53	0.48	0.52	0.47	0.53
LIVE	Opinion that gay men and lesbians should be free to live their own lives as they wish	-0.69	-0.68	-0.9	- 0.82	-0.84
BE	Participation in sports	0.81	0.79	0.77	0.85	0.86
BE	Attendance at ballet, dance, opera	0.69	0.75	0.64	0.85	0.79
BE	Attendance at cinema	0.61	0.56	0.74	0.8	0.74
BE	Attendance at concerts	0.56	0.52	0.39	0.62	0.55
BE	Museums/Galleries	0.79	0.7	0.71	0.91	0.83
BE	Personal use of internet	0.84	0.82	0.69	0.83	0.85
BE	Internet access in households	0.79	0.84	0.76	0.91	0.89
BE	Work-life balance	0.67	0.81	0.82	0.73	0.82
BE	Participation in lifelong learning and training	0.81	0.79	0.76	0.83	0.86

Notes:
1. Pearson correlation coefficients less than 0.4 are not statistically significant at 95%.
2. The Learning to Know dimension does not include Malta, Cyprus and Ireland.
3. The Learning to Live Together dimension does not include Malta and Lithuania.
4. The ELLI-Index does not include Malta, Cyprus, Ireland and Lithuania.

4.4. Drivers in the ELLI-Index

Singular measures of learning

Main drivers of lifelong learning in the EU Member States, in the ELLI-Index are:

- Trust in other people and absence to sexual discriminations in the Learning to Live Together dimension,

- Using internet at work, enterprises providing CVT courses, and Learning new things at work in the Learning to Do dimension,
- Personal use internet, participation in sports, participation in lifelong learning and training and internet access in households in the Learning to Be dimension.

All these measures show more than 0.83 correlation to the ELLI-Index scores. Interestingly, PISA scores and other indicators in the Learning to Know dimension are less influential, taken singularly (<0.74) (see Table 9).

This finding suggests that while organized forms of education (e.g. secondary or tertiary education) are essential to lifelong learning, they do not suffice; vocational training, learning for personal growth and learning for social cohesion are the main drivers of lifelong learning. Such a quantitative assessment reaffirms the "practice engagement theory" (Reder, 1994) and the related "use it or lose it" hypothesis (Krahn and Lowe, 1998). This means that the daily learning-related habits or job-related training of a European citizen could serve to substitute or compensate for a low level of education. On the other hand, the findings could imply that a lack of engagement in learning-related situations at work, at home or in the community could counteract the positive influence of formal education. In other words, formal education may not be able to sustain lifelong learning. These findings are restricted to country level comparisons and more research on the individual level would be worthy the effort.

Learning dimensions

The four learning dimensions of the conceptual framework account for different aspects of learning, yet partially overlapping and not necessarily separable. This is evident in the strong correlations between them, ranging from 0.72 (Learning to Know with Learning to Do) to 0.86 (Learning to Be with Learning to Do or with Learning to Live Together). As a result, all four dimensions have a well balanced impact on the final ELLI-Index scores (see Table 10). Similar conclusions can be drawn using more sophisticated tools, such as the effective weights approach by Stanley and Wang (1968) or the first-order sensitivity measures Saltelli et al. (2008).

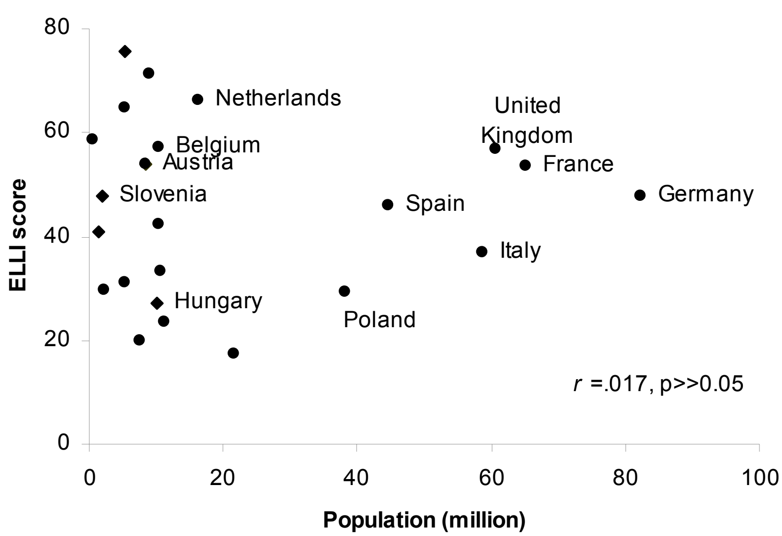
Table 10. Pearson correlation coefficients between the Categories

Dimension	*Be*	*Do*	*Know*	*Live*	*ELLI*	*Explained variance of ELLI scores*
Be	1.00	0.86	0.81	0.86	0.96	93%
Do	0.86	1.00	0.72	0.80	0.92	84%
Know	0.81	0.72	1.00	0.73	0.87	76%
Live	0.86	0.80	0.73	1.00	0.93	87%

4.5 ELLI-Index and population size

The question of whether a certain range of population size can favour development has already been raised (Alesina and Spolaore 2003), but the literature offers no indication of whether population size has an impact on learning. In the EU, the ELLI-Index results show that the highest population sizes –Germany, France, United Kingdom, Italy, Spain, Poland– are associated with moderate to good ELLI-Index scores (Figure 3). Hence none of the most populated EU Member States reach the top five score in lifelong learning in the EU. Although one might have expected that very high population size can favour learning through opportunities and infrastructures, greater community consciousness and citizen engagement, this is not the case in Europe. At lower population sizes (< 20 million) there is no pattern in terms of whether population issues can have a positive or negative impact on lifelong learning. Overall, the association between the Index scores and population is not statistically significant ($r = 0.017, p >> 0.05$), which implies that the ELLI-Index is not biased with respect to population size.

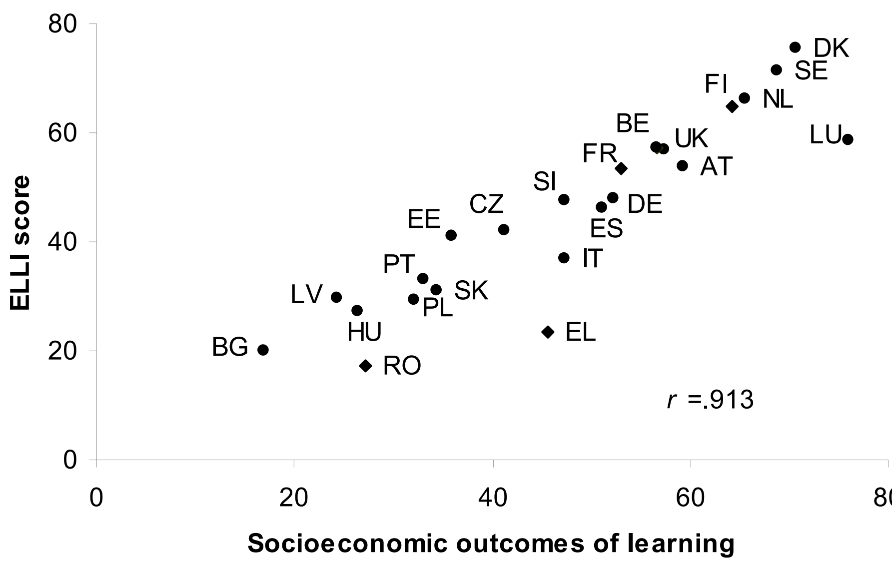
Figure 3. ELLI-Index vs. Population Size in the EU



4.6 ELLI-Index and socioeconomic outcomes of learning

The ELLI-Index was built bearing in mind both the underlying indicators and measures of learning and the economic and social benefits of learning, such as income, employability, population health, life satisfaction, voters' participation and trust in political institutions. These outcomes are often perceived as components of a society's well-being. Figure 4 presents the relationship between the ELLI-Index scores and the aggregate measure of the socioeconomic scores. The results show a high linear relationship between lifelong learning conditions and the economic and social well-being in EU Member States (r = 0.913, n = 23, excluding Malta, Cyprus, Ireland, Lithuania). Countries that have low ELLI-Index scores also have relatively low socioeconomic outcomes scores. Mid-way, Greece despite its relatively moderate to good socioeconomic performance does not perform as high as expected in lifelong learning. Although correlation does not imply causality, and the latter cannot be tested in the EU dataset due to lack of timeseries, these results are consistent with the theory that lifelong learning translates into a more efficient use of an economy's human resources, in terms of employment, civil engagement, and thus affects the overall productivity and economic performance in Europe, in addition to contributing to life satisfaction and personal growth.

Figure 4. ELLI-Index vs. socioeconomic outcomes of learning in the EU



Note: this graph used the socioeconomic outcomes scores rescaled on the same range as the ELLI scores.

Social outcomes of learning. Table 9 shows that the ELLI-Index scores are strongly correlated with material deprivation rate by poverty status in the EU, happiness and life satisfaction, satisfaction with the way democracy works in the home country, trust in political institutions and satisfaction with the home ($r > 0.8$ in absolute terms). Health perception, described by" self-reported health" and "self-reported conditions or health habits" has a moderate to good correlation to the

ELLI-Index scores (around 0.70). Hence, the argument that learning can have a positive influence on people's health by encouraging them to make healthier lifestyle choices and by helping them both to prevent ill health and to manage it when it occurs (Grossman, 2000) is confirmed. Learning is also generally positively associated with social cohesion and democracy (Cambell, 2001). However, Campbell found that learning had both absolute and relative effects on the likelihood of voting and practicing expressive forms of political engagement. The relative effect was weak compared to the absolute effect. Still, this implies that the extent of the absolute effect of learning on voter participation is attenuated by the sorting effects of education. This may in part explain why voting rates and a number of other indicators of social cohesion and democracy have remained stagnant whilst average levels of educational attainment have risen in OECD countries (OECD, 2007), and it might also explain why the ELLI-Index scores do not have a significantly significant association to voter participation rates ($r = .35$) in Europe. However, trust in political institutions and satisfaction with the way democracy works in the home country are strongly associated with the lifelong learning scores in the EU Member States.

Economic outcomes of learning. Human capital theory links education to economic outcomes and offers a robust framework for scientific investigation and policy analysis. More educated people earn more because education increases their productive capacity (Harmon *et al.*, 2003). Highly educated people also have better employment opportunities. The EU reality confirms to a certain extent these relationships, given that the ELLI-Index scores are strongly associated with mean equivalised net income ($r = .83$) and well associated with employment rate ($r = .74$).

5. Uncertainty and sensitivity analysis

The creativity evident in the work of composite indicator developers is not only a response to the multiple demands of the user/stakeholder community but also the result of disagreement within the research community on which indicators influence a particular phenomenon, and by how much. Notwithstanding recent attempts to establish best practice in composite indicator construction (OECD, 2008), "there is no recipe for building composite indicators that is at the same time universally applicable and sufficiently detailed" (Cherchye et al., 2008). This may be due in part to the ambivalent role of composite indicators in both analysis and advocacy (Saltelli, 2007). As the boundaries between the two functions are often blurred, controversy may be unavoidable when discussing these measures.

When building an index to measure lifelong learning in the European Union, it is necessary to take stock of existing methodologies in order to avoid eventual skewness in the assessment and decision-making. By acknowledging the variety of methodological assumptions involved in the development of an index, one can determine whether the main results change substantially when the main assumptions are varied over a reasonable range of possibilities (Saisana *et al.*, 2005; Saltelli et al., 2008). The advantages offered by considering different scenarios to build the Index could be: to gauge the robustness of the Index scores and ranks, to increase its transparency, to identify those countries whose performance improves or deteriorates under certain assumptions, and to help frame the debate on the use of the results for policy making.

The main question to be addressed here is:

- *What models could have been used to build the ELLI-Index and how do the results of these models compare to ELLI?*

We show below how uncertainty analysis (UA) can contribute to such a reflection. UA involves assessing the impact of alternative models on the country ranks. Each model is a different composite indicator in which the choice of normalization, imputation, weights and aggregation method has been varied within a plausible range. This approach helps to avert the criticism frequently dealt to composite measures or rankings, namely that they are presented as if they had been calculated under conditions of certainty (while this is rarely the case) and then taken at face value by end-users (Saisana et al., 2005).

The objective of UA is not to establish the truth or to verify whether the ELLI-Index is a legitimate model, but rather to test whether the country classification and/or its associated inferences are robust or volatile with respect to changes in the methodological assumptions within a plausible and legitimate range. Uncertainty (or robustness) analysis as described by the OECD (2008) has been already used for the assessment of several composite indicators, such as the Multi-dimensional Poverty Assessment Tool (Saisana and Saltelli, 2010), the Composite Learning Index (Saisana, 2008), the Environmental Performance Index (Saisana and Saltelli, 2010), the Alcohol Policy Index (Brand et al., 2007), the Knowledge Economy Index (Saisana and Munda, 2008), the Index of African Governance (Saisana *et al.*,2009) and the University Ranking Systems (Saisana and D'Hombres, 2008).

Furthermore, this part of the analysis aims at identifying those countries for which the ELLI-Index scores/ranks are robust as well as those countries for which it is not. For the first group, policy signals derived from the ELLI-Index can be taken with the confidence that changes in the methodology would have a negligible effect on the country's measured performance. For the latter group, a more cautious approach is advised before translating the ELLI-Index results into policy actions or naming-shaming narratives.

5.1 Multi-modelling approach

A multi-modelling approach was applied in the present work for the purpose of robustness analysis. It consists of exploring, via a saturated sampling, plausible combinations of the main assumptions needed to build the index:

- Four-dimensional structure,
- normalisation method,
- weighting method, and
- aggregation formula.

We identified 25 models, all with their advantages and implications, in order to aggregate the information contained in the ELLI conceptual framework. These models differ in four main aspects: four-dimensional structure (preserved or not), normalisation method (z-scores[6] or Min-

[6] Standardisation (or Z-scores): Each normalised variables value is equal to the raw value minus the average across the EU countries and divided by the standard deviation, so that all normalised variables have similar dispersion across countries. This approach converts all variables to a common scale with an average of zero and standard deviation of one, yet the actual minima and maxima of the standardized values across countries vary among the variables. We standardized, so that each variable has a mean of 50 and a standard deviation of 10, as done by the ELLI-Index developers.

Max[7] approach), weighting method (different statistical methods to derive the weights or equal weighting) and aggregation method (linear, geometric or multi-criteria analysis) (Table 11).

Model 1 resembles the ELLI model, but it differs in that the weights to be assigned to the factor scores within and across the four dimensions do not derive from regression analysis (versus the socioeconomic outcomes). The weights are, instead, estimated as equal to the proportion of the variance explained by a factor, as done for example in the Trade and Development Index (UNCTD, 2005) or in the Summary Indicators of Product Market Regulation (Nicoletti *et al.*, 2000). **Model 2** differs from model 1 in that, instead of z-scores, a Min-max approach is used. In **Model 3 and 4**, we relax the assumption on the four-dimensional structure and let all indicators interact to finally arrive at an overall index using PCA. The two models differ in the normalisation method. In **Model 5 and 6**, PCA is used within each dimension, but all four dimensions are subsequently averaged to produce the overall index score. Again, the two models differ in the normalisation method. The classical equal weighting approach is represented by **Model 7 and 8**, which differ in the normalisation method only. All indicators are simply averaged without considering the four-dimensional structure. In **Model 9 and 10**, we average the indicators within each dimension, and subsequently average the four dimensions. The two models differ in the normalisation method only.

Decision theory practitioners have challenged aggregations based on additive models, such as the ELLI model or models 1-10 above (Eq. 2), because of inherent theoretical inconsistencies (Munda, 2008) and the fully compensatory nature of linear aggregation, in which an x% increase in one indicator can offset an y% decrease in another, where y depends from the ratio of the weights of the two variables. This is the reason why practitioners call weights in linear aggregation 'trade-off coefficients', not to be confused with measures of importance. To account for such challenges we have tried models 11-25. To this end, we applied two alternative

[7] Min-max scaling: Each normalized variable value is equal to the raw value minus the minimum value across countries and divided by the range of values. In this way, the normalized variables have values within [0, 1]. This approach increases the impact of variables with small range of values to the overall Index, but it preserves the information on the different variances between variables. Both these features, depending on the case, could be a desirable or an undesirable property. In our case, the range of values for the variables was set to [1, 100], to allow the use of geometric aggregation which requires strictly positive values.

aggregation functions in those models: either a geometric weighted average (Eq. 3) or a multi-criteria method (Eq. 4)[8].

Weighted Arithmetic Average score: $y_j = \sum_{i=1}^{n} w_i \cdot x_{ij}$ (2)

Weighted Geometric Average score: $y_j = \prod_{i=1}^{n} x_{ij}^{w_i}$ (3)

Borda adjusted score: $y_j = \sum_{i=1}^{n} (m_{ij} + \frac{k_{ij}}{2}) \cdot w_i$ (4)

y_j: composite indicator score for country j, w_i: weight attached to policy category i, x_{ij}: score for country j on policy category i, m_{ij}: number of countries that have weaker performance than country j relative to policy category i; k_{ij}: number of countries with equivalent performance to country j relative to policy category i.

Hence, in **Models 11 to 20**, we employ geometric aggregation, in which the indicators values are raised in a power equal to the weight and subsequently multiplied together into an index. Structure, normalisation and weighting vary as in Models 1-10 where linear aggregation was used. **Models 21 to 25** employ multi-criteria analysis to aggregate the information. Structure and weighting issues vary as previously. Multi-criteria analysis uses ordinal, as opposed to cardinal, information on the indicators values, thus there is no need to normalise the indicators and the raw data are used instead. Finally, Model 26 is based on the cross-efficiency Data Envelopment Analysis approach (further methodological details in Box 1).

[8] Both geometric aggregation and the Borda method applied here are less compensatory than linear weighting. For details see OECD (2008). For an application of the Borda-adjusted method see Brand et al. (2007)

Table 11: 26 models for the development of the ELLI-Index

Model	4-D Structure	Normalisation	Weighting	Aggregation
ELLI	Preserved	z-scores	FA within pillar, Regression weights to Factors, FA pillars, Regression weights to pillars	Linear
M1	Preserved	z-scores	PCA within pillar, PCA pillars	Linear
M2	Preserved	Min-max	PCA within pillar, PCA pillars	Linear
M3	Not preserved	z-scores	PCA all indicators	Linear
M4	Not preserved	Min-max	PCA all indicators	Linear
M5	Preserved	z-scores	PCA within pillar, EW pillars	Linear
M6	Preserved	Min-max	PCA within pillar, EW pillars	Linear
M7	Not preserved	z-scores	EW all indicators	Linear
M8	Not preserved	Min-max	EW all indicators	Linear
M9	Preserved	z-scores	EW within pillar, EW pillars	Linear
M10	Preserved	Min-max	EW within pillar, EW pillars	Linear
M11	Preserved	z-scores	PCA within pillar, PCA pillars	Geometric
M12	Preserved	Min-max	PCA within pillar, PCA pillars	Geometric
M13	Not preserved	z-scores	PCA all indicators	Geometric
M14	Not preserved	Min-max	PCA all indicators	Geometric
M15	Preserved	z-scores	PCA within pillar, EW pillars	Geometric
M16	Preserved	Min-max	PCA within pillar, EW pillars	Geometric
M17	Not preserved	z-scores	EW all indicators	Geometric
M18	Not preserved	Min-max	EW all indicators	Geometric
M19	Preserved	z-scores	EW within pillar, EW pillars	Geometric
M20	Preserved	Min-max	EW within pillar, EW pillars	Geometric
M21	Preserved	Raw data	PCA within pillar, PCA pillars	Multi-criteria
M22	Not preserved	Raw data	PCA all indicators	Multi-criteria
M23	Preserved	Raw data	PCA within pillar, EW pillars	Multi-criteria
M24	Not preserved	Raw data	EW all indicators	Multi-criteria
M25	Preserved	Raw data	EW within pillar, EW pillars	Multi-criteria
M26	Not preserved	z-scores	DEA all indicators	Linear

(EW: Equal weights; FA: Factor Analysis, PCA: Principal Components Analysis, DEA: data envelopment analysis)

Box 1. Data Envelopment Analysis – a candidate model to build the European Lifelong Learning Index

Several policy issues on lifelong learning in Europe entail an intricate balancing act between EU concerns and the country- or region-specific policy priorities. If one opts to compare the multi-dimensional performance of the European Member States by subjecting them to a fixed set of weights, this may prevent acceptance of the index on grounds that a given weighting scheme might not be fair to a particular country. This issue has already been dealt with in Model 26.

In absence of reliable information about the true weights to be attached to the 36 variables underlying the ELLI conceptual framework, we endogenously selected those country-specific weights that maximize a country's score with respect to the EU countries in the dataset using Data Envelopment Analysis (DEA) (Melyn & Moesen, 1991; Cherchye et al., 2004). This gives the following linear programming problem for each country i:

$$Y_i = \max_{wij} \frac{\sum_{j=1}^{14} y_{ij} w_{ij}}{\max_{y_c \in \{dataset\}} \sum_{j=1}^{14} y_{cj} w_{ij}} \qquad (\textit{bounding constraint})$$

Subject to

$w_{ij} \geq 0$, where $j = 1,...,36$, $i = 1,...,23$ (*non-negativity constraint*)

In this basic programming problem, the weights are non-negative and a country's score is between 0 (worst) and 1 (best).

We have also placed reasonable bounds on the weights; otherwise a country could achieve a perfect index score simply by assigning zero weight to those variables for which its performance is very low. To preclude this possibility, we attached upper and lower bounds on the shares, i.e. on the proportion of each variable (a country's score on a given variable multiplied by the variable's weight) over the index score. We requested that the contribution to the overall score of the variables ranges between 1% and 20% at most. Each country is therefore free to decide on the relative contribution of the variables to the overall lifelong learning score, so as to place the country in the best possible position in the ranking, while reflecting the lifelong learning-related priorities of that country. In other words, the DEA method assigns higher contribution to those variables for which a country is strong and a lower weight to those variables for which the country is comparatively weak. However, by assigning these bounds for the shares of the indicators, we ensure that each country includes all the measures/variables and no variable dominates the Index.

However, this traditional DEA model, though suitable for classifying countries into efficient and inefficient ones, it is not very appropriate for ranking countries, since the weights are country-specific. Cross efficiency evaluation method, proposed by Sexton, Silkman, and Hogan (1986), is a DEA extension tool that could be utilized to identify good overall performers and rank countries. The main idea is to use DEA in a peer evaluation instead of a self-evaluation. There are at least three advantages for cross-evaluation method. Firstly, it provides a unique ordering of the countries. Secondly, it eliminates unrealistic weight schemes without requiring the elicitation of weight restrictions from application area experts (Anderson, Hollingsworth, & Inman, 2002). Finally, the cross efficiency means can act effectively to differentiate between good and poor performers (Boussofiane, Dyson, & Thanassoulis, 1991). Therefore the cross-evaluation method is widely used for ranking performance of decision making units (Sexton et al., 1986; Shang & Sueyoshi, 1995).

In brief, the linear programming problem is solved for each country and the n sets of weights are used to calculate n DEA scores for each country. The average of those n scores for each country is used for the overall assessment of countries performance and final ranking.

5.2 Uncertainty analysis results

The results shown in Table 12 are the frequencies of a country's rank in the overall lifelong learning Index calculated across the 26 models. Such a frequency matrix synthesizes the ranking while making the uncertainty explicit. It is beyond doubt that Denmark has the best performance in lifelong learning among the EU Member States (ranks 1 on all 26 models and in the ELLI-Index). Sweden follows the classification (ranks 2nd 84% of the times). Netherlands and Finland follow the ELLI-Index, although given the uncertainties, Finland probably outperforms the Netherlands (Finland is on 3rd rank 53% of the times, whilst the Netherlands occupies the 4th position 53% of the times. A similar swap over in the classification is observed between Luxembourg and Belgium, between United Kingdom and Austria, between Portugal and Slovakia. Similarly for Spain, although it is ranked 12th in the ELLI-Index, ahead of Czech Republic and Estonia, the simulations suggest that Spain underperforms Czech Republic and Estonia in lifelong learning. Similar conclusion is drawn for Spain, whose ELLI-rank places the country ahead of Czech Republic and Estonia, whilst the simulations indicate the opposite. Finally, Portugal and Slovakia outperform Latvia, according to simulations, but the opposite conclusion is drawn according to ELLI. The country profiles in the Annex can give more insight in this respect. Although simulations suggest that some of the methodological choices can have an impact on the ELLI classification of the EU countries, this impact is no more than 2 positions and that overall the ELLI-Index results can be considered as representative of a plurality of models and not just of a specific model.

Summarising the conclusions above, Table 13 shows the median rank and its 99% confidence interval for each EU country across the 26 models. Confidence intervals were estimated using bootstrap (1000 samples taken with replacement, see Efron, 1979). For almost all 23 countries, the ELLI-Index rank falls within this interval, which suggests that these countries were ranked in the correct place, on average. Four countries appear to be slightly misplaced − Estonia, Spain, Latvia and Slovakia, as already discussed above. Any messages drawn on the basis of the ELLI-Index for those four countries should, therefore, be formulated with some caution due to the methodological assumptions made in developing the Index.

A positive result of this analysis is that the narrow confidence interval for all EU countries suggests that there is no particularly volatile section in the graph and that all EU countries see little change in their position, on average (always less than two positions). These narrow confidence intervals suggest that robust conclusions (on average) on the relative performance on lifelong learning of EU countries can be drawn. Furthermore, the high degree of correlation between the ELLI-Index ranking and the simulated median ranking ($r = .99$) produces a high degree of confidence that, for the EU countries, the ELLI-Index classification is reliable and no deliberate bias was introduced in building the ELLI model. In other words, robust narratives can be extracted for almost all countries. Some caution for four countries − Estonia, Spain, Latvia and Slovakia, as already explained.

A question that may arise is whether these four countries, which are moderately sensitive to the methodological assumptions, exhibit the highest variation in their indicator scores. The answer is negative: there is no clear pattern between a country's variability in the 36 indicator scores and the sensitivity of its rank to the methodological assumptions. This implies that there is no means of identifying the countries that are most affected by the methodological choices unless undertaking this type of analysis.

Table 12. ELLI-Index ranks and simulated frequencies across 26 models

	ELLI	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Denmark	1	100																						
Sweden	2		93	7																				
Netherlands	3		7	40	53																			
Finland	4			53	47																			
Luxembourg	5					43	57																	
Belgium	6					53	27	20																
United Kingdom	7							10	63	13	13													
Austria	8					7	13	70	10															
France	9									63	30	7												
Germany	10								27	23	50													
Slovenia	11										7	90	3											
Spain	12											3	13	3	80									
Czech Republic	13												13	67	20									
Estonia	14												70	30										
Italy	15															100								
Latvia	16																7	3	27	30	33			
Portugal	17																13	47	27	13				
Slovakia	18																47	23	23	7				
Poland	19																20	20	17	23	20			
Hungary	20																13	7	7	27	47			
Greece	21																					100		
Bulgaria	22																						100	
Romania	23																							100

Note: Frequencies are calculated across 26 simulated models combining: preserving the four-dimensional structure, normalisation method, weights and aggregation formula. For example, Sweden is never ranked 1st, but instead ranked in the 2nd position in 93% of the simulations (i.e. in 24 out of 26 models), or in the 3rd position only in 7% of the simulations.

Table 13. Simulated median rank and its 99% confidence interval

Country	*ELLI*	*Median*	*Confidence interval for the simulated median rank*
Denmark	1	1	1
Sweden	2	2	2
Netherlands	3	4	[3, 4]
Finland	4	3	[3, 4]
Luxembourg	5	6	[5, 6]
Belgium	6	6	[5, 6]
United Kingdom	7	8	8
Austria	8	7	7
France	9	9	[9, 10]
Germany	10	10	[9,10
Slovenia	11	11	11
Spain	12	14	14
Czech Republic	13	13	13
Estonia	14	12	12, 13]
Italy	15	15	15
Latvia	16	19	[18, 20]
Portugal	17	17	[17, 18]
Slovakia	18	17	[16, 17]
Poland	19	18	[17, 19]
Hungary	20	19	[19, 20]
Greece	21	21	21
Bulgaria	22	22	22
Romania	23	23	23

Note: A country's median rank is calculated over the 26 simulated models generated in our uncertainty analysis. Confidence interval for the median rank are calculated using the bootstrap method with replacement (1000 samples).

5.3 Sensitivity analysis results

Complementary to the uncertainty analysis, a sensitivity analysis makes it possible to assess the impact of a combination of modelling assumptions on the ELLI-Index ranking. Table 14 presents the shifts in rank for the EU countries across the 26 different models with respect to the ELLI-Index rank. In order to assess the overall impact of the different models on the ELLI-Index results, we further summarise the shifts in rank using the median, the maximum and the Spearman rank correlation coefficient, which serve as our sensitivity measures.

The simulations showed that the impact of any of the modelling assumptions is overall negligible: the Spearman rank correlation coefficient between the ELLI-Index ranking and any of the rankings obtained from the various models is greater than 0.97. The vast majority of the EU countries do not see practically any shift in their position, no matter how the modelling assumptions are combined (Belgium, Bulgaria, Czech Rep., Denmark, Greece, Italy, Netherlands, Luxembourg, Poland, Portugal, Romania, Finland, Sweden shift up to one position in the worst/best case). These results imply that all simulated models and the ELLI model point to the same direction in terms of overall country performance in lifelong learning, and that the choice of the model can have a moderate impact (up to four positions shift for Latvia, Hungary and the United Kingdom) for a couple of countries.

This analysis, by assessing the impact of the modelling choices, gives more transparency in the entire process and can help to appreciate the ELLI-Index results with respect to the assumptions made during the development phase. For example, if Hungary wished to improve its overall ELLI-Index rank without even making an effort in improving in some aspects of learning, then it could support the use of model 24 or 25, namely models that employ the Borda-adjusted approach that is less compensatory than the current linear aggregation employed within the ELLI model. Of course this argument could hold for all the countries involved, except for those five countries (Denmark, Italy, Greece, Bulgaria and Romania) that see no shift in their position in the overall classification, no matter how the information in the 36 learning variables is aggregated. For this reason, we resorted, in the previous Section, to presenting the 'median' performance across all models as a summary measure of the plurality of stakeholders' views on how to build a model of lifelong learning.

Table 14. Sensitivity analysis: impact of the modelling assumptions on the ELLI-Index

	ELLI rank	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26
Austria	8	3	2	1	1	3	2	1	1	1	1	2	1	1	0	2	1	1	0	1	0	1	1	1	1	1	3
Belgium	6	-1	-1	0	0	-1	-1	1	1	1	1	-1	0	1	1	-1	0	1	1	1	1	0	1	0	1	1	-1
Bulgaria	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Czech Rep.	13	0	1	-1	-1	0	1	0	0	0	0	1	0	-1	-1	1	0	0	0	0	0	0	-1	0	0	0	1
Denmark	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Estonia	14	2	1	1	1	2	1	2	2	2	2	1	2	1	2	1	2	2	2	2	2	2	1	2	2	2	0
Finland	4	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1
France	9	0	0	0	0	0	0	0	0	0	0	0	-1	0	-1	0	-1	0	-2	0	-2	-1	-1	-1	0	0	1
Germany	10	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	2	1	2	2	2	2
Greece	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hungary	20	0	0	0	0	0	0	3	1	3	1	0	0	0	0	0	0	1	0	1	0	1	2	1	4	4	3
Italy	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Latvia	16	-3	-2	-3	-3	-3	-2	-3	-2	-3	-2	-3	-2	-3	-1	-3	-2	-2	0	-2	0	-4	-4	-4	-4	-4	-3
Luxembourg	5	-1	0	0	0	-1	0	-1	-1	-1	-1	0	0	0	-1	0	0	-1	-1	-1	-1	0	-1	0	-1	-1	-1
Netherlands	3	-1	-1	-1	-1	-1	-1	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	0	-1	0	0	1	0	0	0	-1
Poland	19	1	0	2	1	1	0	-1	-1	-1	-1	1	0	2	0	1	0	-1	0	-1	0	3	3	3	2	2	0
Portugal	17	0	0	1	1	0	0	-1	0	-1	0	0	0	1	1	0	0	0	-1	0	-1	-1	0	-1	-2	-2	0
Romania	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slovakia	18	2	2	0	1	2	2	2	2	2	2	2	2	0	0	2	2	2	1	2	1	1	-1	1	0	0	2
Slovenia	11	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1
Spain	12	-2	-2	0	1	-2	-2	-2	-2	-2	-2	-2	-2	0	-1	-2	-2	-2	-2	-2	-2	-2	0	-2	-2	-2	-1
Sweden	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1
Un.Kingdom	7	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	-1	0	-2	-1	-2	-3	-3	-4
Spearman rank r		0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.97	0.97	0.97

Note: Positive (/negative) numbers represent improvement (/decline) with respect to the ELLI-Index rank. Countries are listed in alphabetical order. Models are described in Table 11.

5.4 Model comparison with respect to the socioeconomic outcomes

The high association between the ELLI-Index ranking and the rankings obtained from the different models discussed above ($r_S > .97$) raises a legitimate question whether the ELLI model that is based on factor regression analysis to estimate the weights produces a more valid composite measure of lifelong learning than the classical equal weighting approach within and across the four dimensions (as in Model 9 or 10). Table 15. provides the answer. Models 9 or 10 are equivalent to the ELLI model with respect to the correlations with the socioeconomic outcomes. The rankings produced by either Model 9 or 10 has the same degree of correlation with the scores on the aggregate socioeconomic outcomes and with other important social or economic outcomes, as the ELLI-model.

This result confirms the expectation that regression weighting will under-perform or perform about as well as equal weighting of standardized indicators (sometimes called 'unit weighting' in the relevant literature) when the coefficient of determination between the model and the external criterion is moderate to low ($R^2 < .50$) (Einhorn and Hogarth, 1975), or when the sample size is small $N \leq 200$ (Raju *et al.*, 1999). Gulliksen (1950) concludes that from a practical point of view, 50-100 indicators are probably sufficient to make differential weighting unprofitable, and the same conclusion is reached if the indicators are strongly correlated. Weighting may be worthwhile, he contends, when there are few (e.g. 3-10 indicators) indicators to be combined and if the average intercorrelation among indicators is also low, e.g. 0.50 or less. In the case of ELLI, the coefficient of determination is high ($R^2 = .80$) but the sample size is small n=23 and the number of underlying measures relatively high (36 measures).

Hence, in order to keep the ELLI-model as simple as possible, this JRC analysis suggests that the classical equal weighting approach within and across the four dimensions could suffice. It has the further advantage that it is easier to communicate to a wide audience[9]. For comparison purposes, Table 16 presents the country ranking in the overall Index and in each of the four learning dimensions under the assumption of equal weighting within and across the four dimensions (raw data were standardised as in the ELLI model) – abbreviated as Model 9 in the analysis above.

[9] Note that the ELLI model was inspired by the CLI model for Canada. However, across the Canadian communities, the 2007 CLI model performed much better than the equal weighting approach with respect to the socioeconomic outcomes of learning (CLI dataset: more than 4500 communities, roughly 20 indicators).

The analysis in this Chapter has taken the conceptual framework for granted. We would argue, though, that a framework mostly reflects the normative assumptions of its developers, and that as such it can be more appropriately the subject of the critique of experts in the field of lifelong learning. The methodological assumptions instead have been tested with the usual tools of applied statistics – by uncertainty and sensitivity analysis.

Table 15. Correlation between the models and the socioeconomic outcomes of learning

	Socioeconomic outcomes	*Mean Equivalised net income*	*Employment rate*	*Self-reported conditions of health or health habits*	*Life satisfaction*	*Gini coefficient*	*Satisfaction with democracy in country*	*Satisfaction at home*
ELLI	0.94	0.90	0.74	0.67	0.91	-0.50	0.84	0.90
M 1	0.93	0.88	0.74	0.67	0.89	-0.56	0.86	0.90
M 2	0.94	0.89	0.74	0.67	0.90	-0.56	0.85	0.91
M 3	0.94	0.89	0.72	0.68	0.91	-0.51	0.86	0.91
M 4	0.94	0.90	0.71	0.68	0.91	-0.52	0.87	0.91
M 5	0.93	0.88	0.74	0.67	0.89	-0.56	0.86	0.90
M 6	0.94	0.89	0.74	0.67	0.90	-0.56	0.85	0.91
M 7	0.92	0.86	0.70	0.66	0.89	-0.60	0.83	0.89
M 8	0.92	0.87	0.73	0.65	0.89	-0.56	0.83	0.90
M 9	0.92	0.86	0.70	0.66	0.89	-0.60	0.83	0.89
M 10	0.92	0.87	0.73	0.65	0.89	-0.56	0.83	0.90
M 11	0.94	0.89	0.73	0.68	0.90	-0.57	0.86	0.91
M 12	0.93	0.88	0.74	0.66	0.91	-0.55	0.86	0.91
M 13	0.94	0.89	0.72	0.67	0.91	-0.52	0.86	0.91
M 14	0.92	0.89	0.75	0.65	0.91	-0.49	0.83	0.89
M 15	0.94	0.89	0.73	0.68	0.90	-0.57	0.86	0.91
M 16	0.93	0.88	0.74	0.66	0.91	-0.55	0.86	0.91
M 17	0.92	0.87	0.73	0.66	0.89	-0.56	0.83	0.90
M 18	0.92	0.87	0.75	0.64	0.91	-0.52	0.83	0.89
M 19	0.92	0.87	0.73	0.66	0.89	-0.56	0.83	0.90
M 20	0.92	0.87	0.75	0.64	0.91	-0.52	0.83	0.89
M 21	0.93	0.87	0.70	0.66	0.92	-0.57	0.87	0.91
M 22	0.92	0.88	0.70	0.67	0.91	-0.52	0.85	0.90
M 23	0.93	0.87	0.70	0.66	0.92	-0.57	0.87	0.91
M 24	0.91	0.84	0.67	0.65	0.90	-0.60	0.84	0.89
M 25	0.91	0.84	0.67	0.65	0.90	-0.60	0.84	0.89
M 26	0.92	0.85	0.71	0.65	0.89	-0.64	0.86	0.91

Note: Spearman rank correlation coefficients less than 0.4 are not statistically significant. Models are described in Table 11.

Table 16. Country rankings in ELLI or under the equal weights assumption

Country	*ELLI – Index rankings*					*Rankings obtained with equal weights within and across the four learning dimensions (Model 9)*				
	Be	**Do**	**Know**	**Live**	**ELLI**	**Be**	**Do**	**Know**	**Live**	**ELLI**
Austria	14	8	16	3	8	13	4	11	1	6
Belgium	9	11	2	8	6	7	11	2	10	7
Bulgaria	27	26	23	20	22	27	24	23	16	22
Cyprus	25	17		14		23	20		8	
Czech Rep.	16	7	20	16	13	16	5	15	22	13
Denmark	1	2	1	1	1	1	1	3	3	1
Estonia	15	15	9	18	14	9	15	7	14	12
Finland	7	3	4	5	4	4	2	1	2	3
France	6	12	8	11	9	8	7	9	15	10
Germany	10	13	14	12	10	10	12	10	9	11
Greece	24	27	22	19	21	25	27	22	18	21
Hungary	22	21	15	25	20	21	21	13	25	19
Ireland	8	10		7		11	6		7	
Italy	17	19	19	13	15	19	18	20	19	16
Latvia	18	25	17	23	18	15	26	16	24	18
Lithuania	19	24	12			18	23	18		
Luxembourg	4	4	10	6	5	5	9	12	4	5
Malta	13	16				14	16			
Netherlands	3	5	6	2	3	3	10	5	6	4
Poland	21	22	13	22	19	22	22	17	20	17
Portugal	23	20	18	15	16	24	19	19	17	20
Romania	26	23	24	24	23	26	25	24	23	23
Slovakia	20	14	21	21	17	20	14	21	21	15
Slovenia	11	6	7	17	11	12	8	8	13	9
Spain	12	18	11	9	12	17	17	14	11	14
Sweden	2	1	3	4	2	2	3	4	5	2
Unit.Kingdom	5	9	5	10	7	6	13	6	12	8

6. Conclusions and policy implications

The ELLI-Index, developed by the *Bertelsmann Foundation* and its international ELLI development team, has the quality features to be used as material for the analysis of lifelong learning in the European Union countries. A high Index (or dimension) score means that a particular country has better lifelong learning performance than a country with much lower scores. While an EU country will score higher than some and lower than others, the purpose of the ELLI-Index is not to identify winners and losers.

The JRC analysis, which was based on the recommendations of the OECD (2008) Handbook on Composite Indicators, suggests that the 2010 ELLI-Index classification provides a reliable picture of the situation at the national level in the EU and can be used to generate a discussion about what policies contribute to lifelong learning, to study the association between lifelong learning and other concepts, such as competitiveness , innovation, and to provide insight into the nature of relevant policy challenges at the EU scale. Besides some fine-tuning issues which were spotted in the ELLI model (see Sections 3 to 5 for detailed discussion and summary in Table 17), this JRC report shows that the ELLI-Index is built according to a sound statistical methodology, its dimensions are well balanced and country ranking's dependence upon input assumptions does not exhibit any of the pathologies which at time affect composite measures. The main recommendation of this study is that the ELLI-model could be simplified into the classical equal weighting approach within and across pillars, without losing any of its quality features.

Table 17. Summary of main recommendations for the ELLI-Index

Data quality issues:
• *Two values in two variables need to be treated prior to applying a linear aggregation in the ELLI model*: For GDP per capita, the value 276.4 for Luxembourg is very high compared to the values for the remaining countries (Ireland's second best value is merely 135.4). Similarly, for "Anyone to discuss intimate and personal matters with", the value 69.2 for Romania is very low compared to the values for the remaining countries (Italy's second low value is 79.5). The recommendation is to winsorize the outlier values by resetting them to the second best/low value (see Table 1).
• *To add a "poor data coverage note" on two variables*: Overall data availability in ELLI is excellent, however, two variables measures - "Involved in work for voluntary or charitable organization" in the Learning to Live Together dimension, and "Satisfaction

with the job" in the socioeconomic outcomes - miss values for almost one-third of the countries. Given that the Index is made of 36+19 variables, eliminating these two variables would leave the results practically unaffected. It is recommended, however, that the two measures are maintained in the conceptual framework but a note on poor data coverage is added.

Structural and modeling issues:

- *A better measure of Environmental consciousness/awareness is needed.* The variable on EPI environment is almost non-significantly correlated with the overall ELLI-Index. This result suggests that if the aim is to include a measure on environment that has an impact on the ELLI-Index, the EPI environment variable in the socio economic outcomes of learning needs to be replaced by another measure that captures better environmental consciousness /awareness.

- *Eventual shift of two measures in different learning dimensions*: The expectation that the variables are more correlated to their conceptual dimension than to any of the other three dimensions of learning is confirmed and furthermore all correlations are statistically significant and have the expected sign (Table 9). There are two exceptions to this expectation worthy of further discussion. First, the measure on "work-life balance" is more correlated to the Learning to Do or Learning to Live Together dimension. Conceptually, the Learning to Live Together dimension captures learning for social cohesion, hence it may be suitable to move the measure on work-life balance from the Learning to Be to the Learning to Live together dimension. Second, the "Labour market expenditure in training" is more correlated to the Learning to Live Together dimension as opposed to the Learning to Do as conceptualised. Hence, labour market spending in training appears to be more related to learning for social cohesion than to vocational learning.

- *Eventually simplify the ELLI model by using equal weights within and across the four dimensions:* The results produced by assuming equal weights within and across the four learning dimensions (with standardised data) are equivalent to those of the ELLI model, as they have the same degree of correlation with the scores on the aggregate socioeconomic outcomes and with other important social or economic outcomes. Hence, in order to keep the ELLI-model as simple as possible, this JRC analysis suggests that the classical equal weighting approach within and across the four dimensions could suffice. It has the further advantage that it is easier to communicate to a wide audience[10].

Dissemination of results:

- Four countries appear to be slightly misplaced – Estonia, Spain, Latvia and Slovakia (see Section 5.2). Any message drawn on the basis of the ELLI-Index for those four countries should, therefore, be formulated with some caution due to the methodological assumptions made in developing the Index.

[10] Note that this result confirms the expectation that regression weighting performs about as well as equal weighting of standardized indicators when the sample size is small (Raju et al., 1999), as in the case of ELLI with just 23 countries. Gulliksen (1950) concludes that from a practical point of view, 50-100 indicators are probably sufficient to make regression weighting unprofitable, and the same conclusion is reached if the indicators are strongly correlated (both assumptions are valid in ELLI, namely there are many variables that are strongly correlated to each other).

References

Alesina, A. and Spolaore E., 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

Aspin, D., Chapman, J., Hatton, M. and Sawano, Y. (Eds.), 2001. *International Handbook of Lifelong Learning*. Dordrecht: Kluwer.

Bandura, R., 2008. *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. UNDP/ODS Working Paper.

Bertelsmann Stiftung, 2010. Making Lifelong Learning Tangible. The European ELLI-Index 2010.

Borda J.C. de, 1784. Mémoire sur les élections au scrutin, in Histoire de l' Académie Royale des Sciences, Paris.

Brand D. A., Saisana M., Rynn L. A., Pennoni F., Lowenfels A. B., 2007. Comparative Analysis of Alcohol Control Policies in 30 Countries. *PLoS Medicine* 4(4): 752-759.

Brennan, R. L. (Ed.), 2006. *Educational measurement*. Westport, CT: American Council on Education/Praeger.

Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O. and Sainsbury, M., 2001. *Framework and specifications for PIRLS assessment*. PIRLS International StudyCenter, Boston: Boston College.

Canadian Council on Learning, 2010. *The 2010 Composite Learning Index: Five Years of Measuring Canada's progress in Lifelong Learning*, Ottawa (Available from http://www.ccl-cca.ca/pdfs/CLI/2010/2010CLI-Booklet_EN.pdf).

Cherchye L., Moesen W., Rogge N., van Puyenbroeck T., Saisana M., Saltelli A., Liska R., Tarantola S., 2008. Creating Composite Indicators with DEA and Robustness Analysis: the case of the Technology Achievement Index. *Journal of Operational Research Society* 59:239-251.

Dave, R. H., 1976. *Foundations of Lifelong Education*. Hamburg: UNESCO Institute of Education.

Delors, J., Al Mufti I., Amagi A., Carneiro R., Chung F., *et al.,* 1996. *Learning: The Treasure Within,* Paris: UNESCO.

Dempster, A.P.; Laird, N.M.; Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society. B 39 (1): 1–38.

Dunteman G.H., 1989. Principal components analysis. Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, No. 69.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7(1), 1–26.

Einhorn, H. and Hogarth, R., 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, **13**, 171-192.

European Civil Society, 2004. *Developing Key Competences: Report of 25 Best Practices*. Brussels: Directorate General of Education and Culture.

Groeneveld, R.A. and Meeden, G., 1984. Measuring skewness and kurtosis. *The Statistician*, 33, 391-399.

Grossmann, M., 2000. The human capital model. In *Handbook of Health Economics Volume 1A* (eds A. J. Culyer, J. P. Newhouse) pp. 347-408. Amsterdam: Elsevier.

Gulliksen, H., 1950. *Theory of Mental Tests*. New York: Wiley.

Harmon, C., Oosterbeek, H., and Walker, I., 2003. The returns to education: microeconomics, *J. Economic Surveys*, **17**, 115-155.

Helland, I.S., 1990. PLS regression and statistical models, *Scand. J. Statist.*, **17**, 97–114.

Jolliffe, I. T., 1982. A note on the use of principal components in regression. *J. R. Statist. Soc.* C, **31**, 300–303.

Kaplan, D., 2000. Structural Equation Modeling: Foundations and Extensions. *SAGE,* Advanced Quantitative Techniques in the Social Sciences series, vol. 10, ISBN 0-7619-1407-2

Krahn, H. and Lowe, G. S., 1998. Literacy utilization in Canadian workplaces. Ottawa: Statistics Canada and Human Resource Development Canada.

Levy, J., 1994. Learning and Foreign Policy: Sweeping a Conceptual Minefield. *International Organization*, **48**, 279–312.

Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with missing data. IInd edition; John Wiley & Sons, Inc.
Manly B., 1994. *Multivariate statistical methods*, Chapman & Hall, UK.
Munda, G., 2008. Social Multi-Criteria Evaluation for a Sustainable Economy. Springer-Verlag: Berlin,Heidelberg.
OECD, 2007. *Understanding the Social Outcomes of Learning*, Paris.
OECD/JRC, 2008. Handbook on Constructing Composite Indicators. Methodology and user Guide, Paris: OECD Publishing (Available from http://composite-indicators. jrc.ec.europa.eu/Handbook.htm).
Raju, N., Bilgic, R., Edwards, J. and Fleer, P., 1999. Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement*, **23**, 99-115.
Reder, S., 1994. Practice-engagement theory: A socio-cultural approach to literacy across languages and cultures. In *Literacy Across Languages and Cultures* (eds B. M. Ferdman, R. M. Weber, A. G. Ramirez) pp. 33–74. Albany: State University of New York Press.
Saisana M., 2008. The 2007 Composite Learning Index: Robustness Issues and Critical Assessment, EUR 23274, European Commission, JRC-IPSC, Ispra, Italy.
Saisana M., D'Hombres B., 2008. Higher Education Rankings: Robustness Issues and Critical Assessment, EUR 23487, European Commission, JRC-IPSC, Ispra, Italy.
Saisana M., Munda G., 2008. Knowledge Economy: measures and drivers, EUR 23486, European Commission, JRC-IPSC, Ispra, Italy.
Saisana M., Saltelli A., 2010. Uncertainty and Sensitivity Analysis of the 2010 Environmental Performance Index, EUR 56990, European Commission, JRC-IPSC, Ispra, Italy.
Saisana M., Tarantola S., Saltelli A., 2005. Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A* 168(2):307-323.
Saisana, M., Annoni, P., Nardo, M., 2009. A robust model to measure governance in African countries, EUR 23773, European Commission, JRC-IPSC, Ispra, Italy.
Saltelli A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, 2008. *Global sensitivity analysis*. The Primer, John Wiley & Sons, England.
Saltelli, A., 2007. Composite indicators between analysis and advocacy. *Social Indicators Research* 81(1), 65-77.
Stanley J.C., Wang M.D., 1968. Differential weighting: a survey of methods and empirical studies. Johns Hopkins Univ., Baltimore.
Stiglitz, J.E., Sen, A., Fitoussi, JP, 2009, Report by the Commission on the Measurement of Economic Performance and Social Progress, www.stiglitz-sen-fitoussi.fr.
Tukey, J.W., 1977. *Exploratory Data Analysis*. MA: Addison-Wesley.

European Commission

EUR 24529 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: ELLI-Index: a sound measure for lifelong learning in the EU
Author(s): Michaela Saisana
Luxembourg: Office for Official Publications of the European Union
2010 – 47 pp. – 21 x 29.70 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-15629-8
doi:10.2788/145

Abstract

The European Lifelong Learning Indicators (ELLI) project is an initiative led by the Bertelsmann Foundation, and one of its aims is to develop, test and pilot a new aggregate measure, the ELLI-Index, for country-level assessment of lifelong learning in the EU Member States. The conceptual framework for the ELLI-Index is loosely based on the UNESCO's International Commission on Education for the Twenty-first Century and the four major dimensions of learning identified: (a) Learning to Know (includes acquisition of knowledge and mastery of learning tools such as concentration, memory and analysis), (b) Learning to Do (concerns occupational, hands-on and practical skills), (c) Learning to Live Together (concerns learning that strengthens cooperation and social cohesion), and (d) Learning to Be (includes the fulfilment of a person, as an individual/member of a family/citizen).

The JRC analysis, which was based on the recommendations of the OECD (2008) Handbook on Composite Indicators, suggests that the 2010 ELLI-Index classification provides a reliable picture of the situation at the national level in the EU and can be used to generate a discussion about what policies contribute to lifelong learning, to study the association between lifelong learning and other concepts, such as competitiveness, innovation, and to provide insight into the nature of relevant policy challenges at the EU scale. Besides some fine-tuning issues which were spotted in the ELLI model (see Sections 3 to 5 for detailed discussion and summary in Table 17), this JRC report shows that the ELLI-Index is built according to a sound statistical methodology, its dimensions are well balanced and country ranking's dependence upon input assumptions does not exhibit any of the pathologies which at time affect composite measures. The main recommendation of this study is that the ELLI-model could be simplified into the classical equal weighting approach within and across pillars, without losing any of its quality features.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

EUROPEAN COMMISSION

LB-NA-24529 -EN-C

Publications Office

ISBN 978-92-79-15629-8