nazioarteko estatistika mintegia euskadin

2009

handbook on constructing composite indicators

adierazle konposatuen garapena eta ebaluazioa

desarrollo y evaluación de indicadores compuestos

STEFANO TARANTOLA
MASSIMILIANO MASCHERINI

50

seminario internacional de estadística en euskadi

2009

handbook on constructing composite indicators

adierazle konposatuen garapena eta ebaluazioa

desarrollo y evaluación de indicadores compuestos

STEFANO TARANTOLA
MASSIMILIANO MASCHERINI

50

Eustat

Lanketa / Elaboración:

Euskal Estatistika Erakundea
Instituto Vasco de Estadística (EUSTAT)

Argitalpena / Edición:

Euskal Estatistika Erakundea
Instituto Vasco de Estadística
Donostia – San Sebastián, 1 – 01010 Vitoria – Gasteiz

Euskal AEko Administrazioa
Administración de la C.A. de Euskadi

Ale-kopurua / Tirada:
500 **ale** / ejemplares

XI-2009

Inprimaketa eta Koadernaketa*:*
Impresión y Encuadernacion:
Estudios Gráficos ZURE S.A.
Ctra. Lutxana-Asua, 24 A
Erandio-Goikoa (BIZKAIA)

I.S.B.N.: 978-84-7749-460-7
Lege-gordailua / Depósito Legal: BI-2948-09

AURKEZPENA

Estatistika nazioarteko mintegiak sustatuz, Eustatek-Euskal Estatistika Erakundeak- zenbait helburu ditu gogoan:

- Unibertsitatearekiko lankidetza sustatzea, bereziki estatistika sailekin.
- Estatistikaz inongo interesik eduki dezaketenen –funtzionario, irakasle, ikasle-birziklatze profesionala erraztea.
- Euskadira estatistika arloko nazioarteko ospe ezaguneko irakasle eta ikertzaileak ekartzea, elkar ezagutzeari eta esperientziak trukatzeari begira.

Ekintza osagarri gisa, ahalik pertsona eta erakunde gehienetara iristeko asmoz eta gaia hobeto ezagutu dadin gure herrian, ikastaro horietako txostenak zabaltzea erabaki da. Testuen zehaztasun teknikoaz arduraturik, argitalpenean txostengileek erabilitako hizkuntzari eutsiko zaio.

Vitoria-Gasteiz, 2009ko azaroa

VÍCTOR URRUTIA ABAIGAR
EUSTATeko Zuzendari Nagusia

PRESENTATION

EUSTAT (Basque Statistics Office) is organising the International Statistics Seminars with the following objectives in mind:

- To foster partnerships with universities and, in particular, with the Departments of Statistics.
- To facilitate the professional recycling of civil servants, lecturers, students and any other parties that may be interested in the field of statistics.
- To bring internationally renowned professors and researchers in the field of statistics to the Basque Country, with the ensuing positive effect of fostering personal relations and exchanging professional experiences.

As a complementary measure, we have decided to disseminate the papers of these courses in order to reach the greatest number of individuals and institutions and thus consolidate knowledge in this area in our country. In order to ensure the greatest technical accuracy of the texts, the papers will be published in their original language.

Vitoria-Gasteiz, November 2009.

VÍCTOR URRUTIA ABAIGAR
EUSTAT General Director

PRESENTACIÓN

Al promover los Seminarios Internacionales de Estadística, EUSTAT –Instituto Vasco de Estadística– pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y de cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a profesores e investigadores de reconocido prestigio internacional en materia estadística, con el consiguiente efecto positivo para el fomento de las relaciones personales y para la transmisión de las experiencias profesionales.

Como actuación complementaria, para llegar al mayor número posible de personas e instituciones y acrecentar así el conocimiento de esta materia en nuestro país, se ha decidido difundir las ponencias de estos cursos. En aras de una mayor precisión técnica de los textos, se mantendrá en su publicación la lengua original de los ponentes.

Vitoria-Gasteiz, noviembre 2009.

VÍCTOR URRUTIA ABAIGAR
Director General de EUSTAT

BIOGRAFI OHARRAK

Stefano Tarantola Europako Batzordearen Ikerketa Zentro Bateratuko (*JRC, Joint Research Centre*) funtzionario zientifikoa da. Ingeniaritzan goi-mailako lizentziatura lortu zuen Milango Unibertsitate Politeknikoan, eta baita Ingeniaritzarako Zientzia eta Teknologiako doktoretza ere. Egun Erkidegoko politiketara bideratutako adierazle estatistikoen eta adierazle konposatuen metodologiaren alorrean egiten du lan. Zientifikoki ebaluatutako saioen egilea da; era berean, egilekide da sentsibilitatea eta ziurgabetasuna aztertzen duten beste lau liburutan, eta baita adierazle konposatuen garapenari buruzko Ekonomiako Lankidetza eta Garapenerako Antolakundearen (ELGA) gidaliburu batean ere.

Massimiliano Mascherini Europako Batzordearen Ikerketa Zentro Bateratuko (*JRC, Joint Research Centre*) funtzionario zientifikoa da. Florentziako Unibertsitateko doktore ere bada Estatistika Aplikatuan. Nazioarteko hainbat estatistika erakundeetako kide da. Sydney-ko Unibertsitatean (Australia) eta Aalborg-eko Unibertsitatean (Danimarka) gonbidatutako irakasletzat aritu izan da. Gehien interesatzen zaion alorra adierazle konposatu eta probabilitate-sareena da. Azken urteotan estatistika aplikatuari eta adierazle konposatuei buruzko argitalpen ugari idatzi ditu nazioarteko aldizkaritan, hala nola *Social Indicator Resarch* eta *Lecture Notes in Artificial Intelligence* argitalpenetan.

BIOGRAPHICAL SKETCH

STEFANO TARANTOLA is a scientific officer at the European Commission, Joint Research Centre with an MSc in Engineering and a PhD in Science and Technologies for Engineering at the Polytechnic of Milan. He conducts methodological work in the field of statistical indicators and composite indicators for EU policy making. He has experience in systems analysis, modelling and in methods to perform robustness analysis of decision processes to policy assumptions. He combines sensitivity analysis and participatory methods for the assessment of the robustness of composite indicators and develops methodologies for sensitivity analysis. He is the author of papers in the peer-reviewed literature, co-author of four books on uncertainty and sensitivity analysis and of a handbook on composite indicators development with the OECD.

MASSIMILIANO MASCHERINI is a scientific officer at the European Commission, Joint Research Centre with a PhD in Applied Statistics at the University of Florence. He is a member of several international statistical societies. He has been visiting fellow at the University of Sydney (Australia) and at Aalborg University (Denmark). His main research interests focus on composite indicators and probabilistic networks. Over the last years he produced a steady flow of publications on applied statistics and composite indicators in international journals such as Social Indicator Research and Lecture Notes in Artificial Intelligence.

NOTAS BIOGRÁFICAS

STEFANO TARANTOLA es funcionario científico del Centro Común de Investigación de la Comisión Europea (Joint Reserach Centre, JRC). Es Licenciado Superior en Ingeniería y Doctor en Ciencia y Tecnologías para la Ingeniería por la Universidad Politécnica de Milán. En la actualidad, trabaja en el campo de la metodología de los indicadores estadísticos y los indicadores compuestos para la realización de políticas comunitarias. Cuenta con una gran experiencia en análisis de sistemas y modelización y, además, en métodos de aplicación de análisis robustos a la toma de decisiones políticas. Combina análisis de sensibilidad y métodos participativos para la valoración de la solidez de indicadores compuestos, desarrollando metodologías para el análisis de sensibilidad. Es autor de ensayos y trabajos dentro de un comité evaluador; co-autor de cuatro libros sobre análisis de sensibilidad e incertidumbre, y de un manual sobre desarrollo de indicadores compuestos con la Organización para la Cooperación y Desarrollo Económicos (OECD).

MASSIMILIANO MASCHERINI es funcionario científico del Centro Común de Investigación de la Comisión Europea (Joint Reserach Centre, JRC) y doctor en Estadística Aplicada por la Universidad de Florencia. Pertenece a varias sociedades estadísticas internacionales. Ha sido profesor invitado en la Universidad de Sydney (Australia) y en la Universidad Aalborg (Dinamarca). Su principal campo de interés son los indicadores compuestos y las redes de probabilidades. Durante los últimos años ha producido un flujo continuado de publicaciones sobre estadística aplicada e indicadores compuestos en revistas internacionales, como el *Social Indicator Research* y el *Lecture Notes in Artificial Intelligence*.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

This booklet is composed of three sections. The first section, based on Nardo et al. 2008, suggests the basic ten steps for the development of composite indicators, from the development of the theoretical framework to the presentation and dissemination of the results. The second section presents the active citizenship composite indicator that is built following the guidelines described in the first section. The active citizenship composite indicator is part of the “Active Citizenship for Democracy” research project, coordinated by the Centre for Research on Education and Lifelong Learning (CRELL) in cooperation with the Directorate General Education and Culture of the European Commission and the Council of Europe. In the final section, the ongoing debate on composite indicators is addressed, highlighting the importance of investigating the robustness of the results to a variety of uncertain factors.

2. Steps for developing composite indicators

An indicator is a statistical construct aimed at describing what is happening in nature, economy, society, etc.

It is obtained using a combination of raw data often in a modeling exercise. A well- known definition of *an indicator* in the literature is <<*something that provides a clue to a matter of larger significance or makes perceptible a trend or phenomenon that is not immediately detectable*>> (Hammond et al, 1995). An indicator's main defining characteristic is that it quantifies and simplifies information in a manner that promotes the understanding of the topic to both decision-makers and the public.

Composite indicators are mathematical combinations of underlying indicators that have no common meaningful unit of measurement (Saisana and Tarantola, 2002). In the US there is general preference to call them *indices*.

Composite indicators are very common in fields such as economic and business statistics, and in a variety of policy domains such as industrial competitiveness, sustainable development, globalisation and innovation (Saltelli, 2007).

The proliferation of this kind of indicators is a clear symptom of their political importance and operational relevance in decision making (Nardo et al., 2008). To give an idea, a search on “Scholar Google” for the word ‘composite indicators’ in July 2009 returns 3,870 hits. In October 2005 these were 992.

The objective of composite indicators is to measure multi-dimensional concepts that cannot be captured by a single indicator. Ideally, a composite indicator should be based on:

- a solid theoretical framework (structure of the phenomenon),
- a sound process of construction (indicators selection and aggregation)
- and good quality underlying data.

A good technical preparation can make the composite indicator more robust to uncertainties, more resilient (in the sense that it remains relevant over time) in a way that facilitates negotiation rather than exacerbating disagreement. However, good technical preparation with poor underlying data will produce poor results. That is why the three aspects represent necessary conditions in order to achieve solid and defensible composite indicators in the dialogue with stakeholders. In other words, transparency and coherence should guide the entire process, from the framing of the problem to the interpretation of results.

Nardo et al. 2008 suggest ten steps for the development of composite indicators, from the development of a theoretical framework to the presentation and dissemination of a composite indicator. <<*Each step is extremely important, but coherence in the whole process is equally vital. Choices made in one step can have important implications for others: therefore, the composite indicator builder has not only to make the most appropriate methodological choices in each step, but also to identify whether they fit together well.*>>

The rest of this section discusses each of the ten steps and is borrowed from Nardo et al. 2008.

2.1. STEP 1: Developing a theoretical framework

What is badly defined is likely to be badly measured

A sound theoretical framework is the starting point in constructing composite indicators. The framework should clearly define the phenomenon to be measured and its sub-components, selecting individual indicators and weights that reflect their relative importance and the dimensions of the overall composite. This process should ideally be based on what is desirable to measure and not on which indicators are available.

For example, gross domestic product (GDP) measures the total value of goods and services produced in a given country, where the weights are estimated based on economic theory and reflect the relative price of goods and services. The theoretical and statistical frameworks to measure GDP have been developed over the last 50 years and a revision of the 1993 System of National Accounts is currently being undertaken by the major international organizations. However, not all multi-dimensional concepts have such solid theoretical and empirical underpinnings. Composite indicators in newly emerging policy areas, *e.g.* competitiveness, sustainable development, e-business readiness, *etc.*, might be very subjective, since the economic research in these fields is still being developed. Transparency is thus essential in constructing credible indicators. This entails:

- Defining the concept. The definition should give the reader a clear sense of what is being measured by the composite indicator. It should refer to the theoretical framework, linking various sub-groups and the underlying indicators. For example, the Growth Competitiveness Index (GCI) developed by the World Economic Forum is founded on the idea "that the process of economic growth can be analysed within three important broad categories: the macroeconomic environment, the quality of public institutions, and technology." The GCI has, therefore, a clear link between the framework (whatever this is) and the structure of the composite indicator. Some complex concepts, however, are difficult to define and measure precisely or may be subject to controversy among stakeholders. Ultimately, the users of composite indicators should assess their quality and relevance.

- Determining domains, or pillars. Multi-dimensional concepts can be divided into several domains, or pillars. These need not be (statistically) independent of each other, and existing linkages should be described theoretically or empirically to the greatest extent possible. The Technology Achievement Index, for example, is conceptually divided into four pillars of technological capacity: creation of technology, diffusion of recent innovations, diffusion of old innovations and human skills. Such a nested structure improves the user's understanding of the driving forces behind the composite indicator. It may also make it easier to determine the relative weights across different factors. This step, as well as the next, should involve experts and stakeholders as much as possible, in order to take into account multiple viewpoints and to increase the robustness of the conceptual framework and set of indicators.

- Identifying the selection criteria for the underlying indicators. The selection criteria should work as a guide to whether an indicator should be included or not in the overall composite indicator. It should be as precise as possible and should describe the phenomenon being measured, *i.e.* input, output or process. Too often composite indicators include both input and output measures. For example, an Innovation Index could combine R&D expenditures (inputs) and the number of new products and services (outputs) in order to measure the scope of innovative activity in a given country. However, only the latter set of output indicators should be included (or expressed in terms of output per unit of input) if the index is intended to measure innovation performance.

An interesting example of theoretical framework is the Composite Learning Index developed by the Canadian Council on Learning, which is composed by the four pillars from Jacques Delors' UNESCO Task Force:

Learning to Know: involving developing the foundation of skills and knowledge needed to function in the world. This includes literacy, numeracy, general knowledge and critical thinking.

Learning to Do: referring to the acquisition of applied skills. It can encompass technical and hands-on skills and knowledge, and is closely tied to occupational success.

Learning to Live Together: involving developing values of respect and concern for others, fostering social and inter-personal skills, and an appreciation of the diversity of Canadians. This area of learning contributes to a cohesive society.

Learning to Be: referirng to the learning that helps develop the whole person-mind, body and spirit. This aspect concerns personal discovery, self-knowledge, creativity and achieving a healthy balance in life.

Other interesting examples are the theoretical framework for the Summary Innovation Index 2005 for the period 2005-2007, consisting of four levels (the composite, the input/output domains, the 5 sub-domains *drivers, knowledge creation, innovation & entrepreneurship, application, intellectual property,* and the 25 underlying indicators) and its revision for the period 2008-2010.

By the end of Step 1 the developer should have:

- A clear understanding and definition of the multi-dimensional phenomenon to be measured.
- A nested structure of the various sub-groups of the phenomenon if needed.
- A list of selection criteria for the underlying variables, *e.g.* input, output, process.
- Clear documentation of the above.

2.2. STEP 2: Selecting variables

A composite indicator is above all the sum of its parts

The strengths and weaknesses of composite indicators largely derive from the quality of the underlying variables. Ideally, variables should be selected on the basis of their relevance, analytical soundness, timeliness, accessibility, *etc*.

The theoretical framework should guide the choice of the underlying indicators. The selection process can be quite subjective and therefore should involve stakeholders.

For certain sectors the scarcity of internationally comparable quantitative (hard) data could limit the developer's ability to build a good-quality composite indicator. Qualitative (soft) data from surveys or policy reviews are often used and proxy indicators considered. These latter can be used when the desired data are unavailable or when cross-country comparability is limited. For example, data on the number of employees that use computers might not be available. Instead, the number of employees who have access to computers could be used as a proxy. As in the case of soft data, caution must be taken in using proxy indicators. To the extent that data permit, the accuracy of proxy measures should be checked through correlation and sensitivity analysis. The builder should also pay close attention to whether the indicator in question is dependent on GDP or other size-related factors. To have an objective comparison across small and large countries, scaling of variables by an appropriate size measure, *e.g.* population, income, trade volume, and populated land area, *etc*. is required. Finally, the type of variables selected - input, output or process indicators - must match the definition of the intended composite indicator.

The quality and accuracy of composite indicators should evolve in parallel with improvements in data collection and indicator development. The current trend towards developing composite indicators of country performance in a range of policy areas may provide further impetus to improving data collection, identifying new data sources and enhancing the international comparability of statistics. On the other hand we do not marry the idea that using what is available is necessarily enough. Poor data will produce poor results in a "garbage-in, garbage-out logic. From a pragmatic point of view, however, compromises need to be done when constructing a composite. What we deem essential is the transparency of these compromises.

A good practice for selecting indicators, based on five criteria, is suggested in the Summary Innovation Index of the European Commission's DG Enterprise and Industry:

- **Policy relevance:** with the aim of identifying indicators that are meaningful for decisional processes and reflective of the political orientations (i.e. Lisbon objectives);

- **Redundancy:** when 2 indicators are found to be redundant, which means that they give the same information, it is recommended to select only one.

- **First comer privilege:** when two indicators are redundant, it is recommended to select the one that was already included in the previous scoreboard.

- **Correlation:** when 2 indicators are highly correlated and convey strong political messages, they can be both included in the final list.

- **Availability:** indicators which prove to be available for a large number of countries, and which can be extracted from regularly updated databases are recommended.

By the end of Step 2 the developer should have:

- Checked the quality of the available indicators.
- Discussed the strengths and weaknesses of each selected indicator.
- Created a summary table on data characteristics, *e.g.* availability (across country, time), source, type (hard, soft or input, output, process).

2.3. STEP 3: Imputation of missing data

The idea of imputation could be both seductive and dangerous

It is very rare that our dataset (i.e. the matrix indicators x countries) is complete. Missing data often hinder the development of robust composite indicators. Data can be missing in a random or non-random fashion. The missing patterns could be:

Missing completely at random (MCAR). Missing values do not depend on the variable of interest or on any other observed variable in the data set. For example, the missing values in variable income would be of the MCAR type if (i) people who do not report their income have, on average, the same income as people who do report income; and if (ii) each of the other variables in the data set would have to be the same, on average, for the people who did not report their income and the people who did report their income.

Missing at random (MAR). Missing values do not depend on the variable of interest, but are conditional on other variables in the data set. For example, the missing values in income would be MAR if the probability of missing data on income depends on marital status but, within each category of marital status, the probability of missing income is unrelated to the value of income. Missing by design, *e.g.* if survey question 1 is answered yes, then survey question 2 is not to be answered, are also MAR as missingness depends on the covariates.

Not missing at random (NMAR). Missing values depend on the values themselves. For example, high income households are less likely to report their income.

Unfortunately, there is no statistical test for NMAR and often no basis on which to judge whether data are missing at random or systematically, while most of the methods that impute missing values require a missing at random mechanism, *i.e.* MCAR or MAR. When there are reasons to assume a non-random missing pattern (NMAR), the pattern must be explicitly modelled and included in the analysis. This could be very difficult and could imply *ad hoc* assumptions that are likely to influence the result of the entire exercise.

There are three general methods for dealing with missing data: (i) case deletion, (ii) single imputation or (iii) multiple imputation. The first option simply omits the missing records from the analysis. For composite indicators this means either omitting an indicator for all countries, or omitting one country implying discarding other 'expensive-to-collect' data. This option is in general not practicable as we wish to provide scores for all countries keeping the maximum number of underlying indicators. In addition, this approach ignores possible systematic differences between complete and incomplete samples and produces unbiased estimates only if deleted records are a random sub-sample of the original sample (MCAR assumption). Furthermore, standard errors will generally be larger in a reduced sample, given that less information is used. As a rule of thumb, if a variable has more than 5% missing values, cases are not deleted (Little & Rubin, 2002).

The other two approaches consider the missing data as part of the analysis and try to impute values through either single imputation, *e.g.* mean/median/mode substitution, regression, hot- and cold-deck, expectation-maximization, or multiple imputation, *e.g*. Markov Chain Monte Carlo algorithm. Data imputation could lead to the minimization of bias and the use of 'expensive to collect' data that would otherwise be discarded by case deletion. However, it can also allow data to influence the type of imputation. In the words of Dempster & Rubin (1983):

> The idea of imputation is both seductive and dangerous. It is seductive because it can lull the user into the pleasurable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to real and imputed data have substantial bias.

No imputation model is free of assumptions and the imputation results should hence be thoroughly checked for their statistical properties, such as distributional characteristics, as well as heuristically for their meaningfulness, *e.g* whether negative imputed values are possible.

With imputation, uncertainty analysis becomes important as the uncertainty in the imputed data should be acknowledged in the country scores, making possible to take into account the effects of imputation in the course of the analysis.

By the end of Step 3 the constructor should have:

- A complete data set without missing values.
- A measure of the reliability of each imputed value so as to explore the impact of imputation uncertainty on the composite indicator scores.
- Discussed the presence of outliers in the dataset.
- Documented and explained the selected imputation procedures and the results.

2.4. STEP 4: Multivariate analysis

Analyzing the underlying structure of the data is still an art

Over the last few decades, there has been an increase in the number of composite indicators being developed by various national and international agencies. Unfortunately, individual indicators are sometimes selected in an arbitrary manner with little attention paid to the interrelationships between them. This can lead to composites which overwhelm, confuse and mislead decision-makers and the general public. Some analysts characterize this environment as "indicator rich but information poor". The underlying nature of the data needs to be carefully analyzed before the construction of a composite indicator. This preliminary step is helpful in assessing the suitability of the data set and will provide an understanding of the implications of the methodological choices, *e.g.* weighting and aggregation, during the development phase of the composite indicator. Information can be grouped and analyzed along at least two dimensions of the data set: individual indicators and countries.

- **Grouping information on individual indicators**. The analyst must first decide whether the structure of the composite indicator is well defined (see Step 1) and whether the set of available individual indicators is sufficient or appropriate to describe the phenomenon (see Step 2). This decision can be based on expert opinion and the statistical structure of the data set. Different analytical approaches, such as principal components analysis, can be used to explore whether the dimensions of the phenomenon are statistically well-balanced in the composite indicator. If not, a revision of the individual indicators might be necessary.

  The goal of principal components analysis (PCA) is to reveal how different variables change in relation to each other and how they are associated. This is achieved by transforming correlated variables into a new set of uncorrelated variables using a covariance matrix or its standardised form - the correlation matrix. Factor analysis (FA) is similar to PCA, but is based on a particular statistical model. An alternative way to investigate the degree of correlation among a set of variables is to use the Cronbach coefficient alpha (c-alpha), which is the most common estimate of internal consistency of items in a model or survey. These multivariate analysis techniques are useful for gaining insight into the structure of the data set of the composite. However, it is important to avoid carrying out multivariate analysis if the sample is small compared to the number of indicators, since results will not have known statistical properties.

- **Grouping information on countries**. Cluster analysis is another tool for classifying large amounts of information into manageable sets. It has been applied to a wide variety of research problems and fields. Cluster analysis is also used in the development of composite indicators to group information on countries based on their similarity on different underlying indicators. Cluster analysis serves as: (i) a purely statistical method of aggregation of the indicators, ii) a diagnostic tool for exploring the impact of the methodological choices made during the construction phase of the composite indicator, (iii) a method of disseminating information on the composite indicator without losing that on the dimensions of the individual indicators, and (iv) a method for selecting groups of countries for the imputation of missing data with a view to decreasing the variance of the imputed values.

When the number of variables is large or when it is believed that some of them do not contribute to identifying the clustering structure in the data set, continuous and discrete models can be applied sequentially. Researchers frequently carry out a PCA and then apply a clustering algorithm on the object scores on the first few components, called “tandem analysis”. However, caution is required, as PCA or FA may identify dimensions that do not necessarily help to reveal the clustering structure in the data and may actually mask the taxonomic information (Table 1).

Various alternative methods combining cluster analysis and the search for a low-dimensional representation have been proposed, focusing on multi-dimensional scaling or unfolding analysis. Factorial k-means analysis combines k-means cluster analysis with aspects of FA and PCA. A discrete clustering model together with a continuous factorial model are fitted simultaneously to two-way data to identify the best partition of the objects, described by the best orthogonal linear combinations of the variables (factors) according to the least-squares criterion. This has a wide range of applications since it achieves a double objective: data reduction and synthesis, simultaneously in the direction of objects and variables. Originally applied to short-term macroeconomic data, factorial k-means analysis has a fast alternating least-squares algorithm that extends its application to large data sets. This methodology can be recommended as an alternative to the widely-used tandem analysis.

Table 1. Strengths and weaknesses of multivariate analysis

	Strengths	Weaknesses
Principal Components/ Factor Analysis	▪ Can summarise a set of individual indicators while preserving the maximum possible proportion of the total variation in the original data set. ▪ Largest factor loadings are assigned to the individual indicators that have the largest variation across countries, a desirable property for cross-country comparisons, as individual indicators that are similar across countries are of little interest and cannot possibly explain differences in performance.	▪ Correlations do not necessarily represent the real influence of the individual indicators on the phenomenon being measured. ▪ Sensitive to modifications in the basic data: data revisions and updates, *e.g.* new countries. ▪ Sensitive to the presence of outliers, which may introduce a spurious variability in the data. ▪ Sensitive to small-sample problems, which are particularly relevant when the focus is on a limited set of countries. ▪ Minimisation of the contribution of individual indicators which do not move with other individual indicators.
Cronbach Coefficient Alpha	▪ Measures the internal consistency in the set of individual indicators, *i.e.* how well they describe a uni-dimensional construct. Thus it is useful to cluster similar objects.	▪ Correlations do not necessarily represent the real influence of the individual indicators on the phenomenon expressed by the composite indicator. ▪ Meaningful only when the composite indicator is computed as a 'scale' (*i.e.* as the sum of the individual indicators).
Cluster Analysis	▪ Offers a different way to group countries; gives some insight into the structure of the data set.	▪ Purely a descriptive tool; may not be transparent if the methodological choices made during the analysis are not motivated and clearly explained.

By the end of Step 4 the constructor should have:

- Checked the underlying structure of the data along various dimensions, *i.e.* individual indicators, countries.
- Applied the suitable multivariate methodology, *e.g.* PCA, FA, cluster analysis.
- Identified sub-groups of indicators or groups of countries that are statistically "similar".
- Analysed the structure of the data set and compared this to the theoretical framework.
- Documented the results of the multivariate analysis and the interpretation of the components and factors.

2.5. STEP 5: Normalisation of data

Avoid adding up apples and oranges

Normalisation is required prior to any data aggregation as the indicators in a data set often have different measurement units. A number of normalisation methods exist (Table 2) (Freudenberg, 2003; Jacobs *et al.*, 2004):

1. *Ranking* is the simplest normalisation technique. This method is not affected by outliers and allows the performance of countries to be followed over time in terms of relative positions (rankings). Country performance in absolute terms however cannot be evaluated as information on levels is lost. Some examples that use ranking include: the Information and Communications Technology Index (Fagerberg, 2001) and the Medicare Study on Healthcare Performance across the United States (Jencks *et al*., 2003).

2. *Standardisation* (or z-scores) converts indicators to a common scale with a mean of zero and standard deviation of one. Indicators with extreme values thus have a greater effect on the composite indicator. This might not be desirable if the intention is to reward exceptional behaviour, *i.e.*, if an extremely good result on a few indicators is thought to be better than a lot of average scores. This effect can be corrected in the aggregation methodology, *e.g.* by excluding the best and worst individual indicator scores from inclusion in the index or by assigning differential weights based on the "desirability" of the individual indicator scores.

3. Min-Max normalises indicators to have an identical range [0, 1] by subtracting the minimum value and dividing by the range of the indicator values. However, extreme values/or outliers could distort the transformed indicator. On the other hand, Min-Max normalisation could widen the range of indicators lying within a small interval, increasing the effect on the composite indicator more than the z-score transformation.

4. Distance to a reference measures the relative position of a given indicator vis-à-vis a reference point. This could be a target to be reached in a given time frame. For example, the Kyoto Protocol has established an 8% reduction target for CO2 emissions by 2010 for European Union members. The reference could also be an external benchmark country. For example, the United States and Japan are often used as benchmarks for the composite indicators built in the framework of the EU Lisbon agenda. Alternatively, the reference country could be the average country of the group and would be assigned a value of 1, while other countries would receive scores depending on their distance from the average. Hence, standardised

indicators that are higher than 1 indicate countries with above-average performance. The reference country could also be the group leader, in which the leading country receives 1 and the others are given percentage points away from the leader. This approach, however, is based on extreme values which could be unreliable outliers.

5. Categorical scale assigns a score for each indicator. Categories can be numerical, such as one, two or three stars, or qualitative, such as ‘fully achieved’, ‘partly achieved’ or ‘not achieved’. Often, the scores are based on the percentiles of the distribution of the indicator across countries. For example, the top 5% receive a score of 100, the units between the 85th and 95th percentiles receive 80 points, the values between the 65th and the 85th percentiles receive 60 points, all the way to 0 points, thereby rewarding the best performing countries and penalising the worst. Since the same percentile transformation is used for different years, any change in the definition of the indicator over time will not affect the transformed variable. However, it is difficult to follow increases over time. Categorical scales exclude large amounts of information about the variance of the transformed indicators. Besides, when there is little variation within the original scores, the percentile bands force the categorisation on the data, irrespective of the underlying distribution. A possible solution is to adjust the percentile brackets across the individual indicators in order to obtain transformed categorical variables with almost normal distributions.

6. Indicators above or below the mean are transformed such that values around the mean receive 0, whereas those above/below a certain threshold receive 1 and -1 respectively, *e.g.* the Summary Innovation Index (European Commission, 2001a). This normalisation method is simple and is not affected by outliers. However, the arbitrariness of the threshold level and the omission of absolute level information are often criticised. For example, if the value of a given indicator for country A is 3 times (300%) above the mean, and the value for country B is 25% above the mean, both countries would be counted as ‘above average’ with a threshold of 20% around the mean.

7. Methods for cyclical indicators. The results of business tendency surveys are usually combined into composite indicators to reduce the risk of false signals, and to better forecast cycles in economic activities (Nilsson, 2000). See, for example, the OECD composite leading indicators, and the EU economic sentiment indicators (European Commission, 2004). This method implicitly gives less weight to the more irregular series in the cyclical movement of the composite indicator, unless some prior *ad hoc* smoothing is performed.

8. The latter is a special case of balance of opinions, in which managers of firms from different sectors and of varying sizes are asked to express their opinion on their firm's performance.

9. Percentage of annual differences over consecutive years represents the percentage growth with respect to the previous year instead of the absolute level. The transformation can be used only when the indicators are available for a number of years, *e.g.* Internal Market Index (European Commission, 2001b; Tarantola *et al.*, 2002; Tarantola *et al.*, 2004).

Table 2. Normalisation methods

Method	Equation
1. Ranking	$I_{qc}^{t} = Rank(x_{qc}^{t})$
2. Standardisation (or z-scores)	$I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\bar{c}}^{t}}{\sigma_{qc=\bar{c}}^{t}}$
3. Min-Max	$I_{qc}^{t} = \frac{x_{qc}^{t} - \min_c(x_q^{t_0})}{\max_c(x_q^{t_0}) - \min_c(x_q^{t_0})}$.
4. Distance to a reference country	$I_{qc}^{t} = \frac{x_{qc}^{t}}{x_{qc=\bar{c}}^{t_0}}$ or $I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc=\bar{c}}^{t_0}}{x_{qc=\bar{c}}^{t_0}}$
5. Categorical scales	Example: $I_{qc}^{t} = \begin{cases} 0 & \text{if } x_{qc}^{t} < P^{15} \\ 20 & \text{if } P^{15} \le x_{qc}^{t} < P^{25} \\ 40 & \text{if } P^{25} \le x_{qc}^{t} < P^{65} \\ 60 & \text{if } P^{65} \le x_{qc}^{t} < P^{85} \\ 80 & \text{if } P^{85} \le x_{qc}^{t} < P^{95} \\ 100 & \text{if } P^{95} \le x_{qc}^{t} \end{cases}$
6. Indicators above or below the mean	$I_{qc}^{t} = \begin{cases} 1 & \text{if } w > (1+p) \\ 0 & \text{if } (1-p) \le w \le (1+p) \\ -1 & \text{if } w < (1-p) \end{cases}$ where $w = x_{qc}^{t} / x_{qc=\bar{c}}^{t_0}$
7. Cyclical indicators (OECD)	$I_{qc}^{t} = \frac{x_{qc}^{t} - E_t(x_{qc}^{t})}{E_t(\lvert x_{qc}^{t} - E_t(x_{qc}^{t})\rvert)}$
8. Balance of opinions (EC)	$I_{qc}^{t} = \frac{100}{N_e}\sum_{e}^{N_e} \operatorname{sgn}_e(x_{qc}^{t} - x_{qc}^{t-1})$
9. Percentage of annual differences over consecutive years	$I_{qc}^{t} = \frac{x_{qc}^{t} - x_{qc}^{t-1}}{x_{qc}^{t}}$

Note: x_{qc}^{t} is the value of indicator q for country c at time t. $\bar{c}$ is the reference country. The operator *sgn* gives the sign of the argument (*i.e.* +1 if the argument is positive, -1 if the argument is negative). N_e is the total number of experts surveyed. P^i is the i-th percentile of the distribution of the indicator x_{qc}^{t} and *p* an arbitrary threshold around the mean.

The selection of a suitable method, however, is not trivial and deserves special attention to eventual scale adjustments (Ebert & Welsh, 2004) or transformation or highly skewed indicators. The normalization method should take into account the data properties, as well as the objectives of the composite indicator. Robustness tests might be needed to assess their impact on the outcomes.

By the end of Step 5 the constructor should have:

- Selected the appropriate normalisation procedure(s) with reference to the theoretical framework and to the properties of the data.
- Made scale adjustments, if necessary.
- Transformed highly skewed indicators, if necessary.
- Documented and explained the selected normalisation procedure and the results.

2.6. STEP 6: Weighting and aggregation

The relative importance of the indicators is a source of contention

When used in a benchmarking framework, weights can have a significant effect on the overall composite indicator and the country rankings. A number of weighting techniques exist (Table 3). Some are derived from statistical models, such as factor analysis, data envelopment analysis and unobserved components models (UCM), or from participatory methods like budget allocation processes (BAP), analytic hierarchy processes (AHP) and conjoint analysis (CA). Regardless of which method is used, weights are essentially value judgments. While some analysts might choose weights based only on statistical methods, others might reward (or punish) components that are deemed more (or less) influential, depending on expert opinion, to better reflect policy priorities or theoretical factors.

Most composite indicators rely on equal weighting (EW), *i.e.* all variables are given the same weight. This essentially implies that all variables are "worth" the same in the composite, but it could also disguise the absence of a statistical or an empirical basis, *e.g.* when there is insufficient knowledge of causal relationships or a lack of consensus on the alternative. In any case, equal weighting does not mean "no weights", but implicitly implies that the weights are equal. Moreover, if indicators are grouped into dimensions and those are further aggregated into the composite, then applying equal weighting to the indicators may imply an unequal weighting of the dimension (if the dimensions have different number of indicators), resulting in an unbalanced structure in the composite indicator.

Table 3. Compatibility between aggregation and weighting methods

Weighting methods	Aggregation methods		
	Linear[4]	Geometric[4]	Multi-criteria
EW	Yes	Yes	Yes
PCA/FA	Yes	Yes	Yes
BOD	Yes[1]	No[2]	No[2]
UCM	Yes	No[2]	No[2]
BAP	Yes	Yes	Yes
AHP	Yes	Yes	No[3]
CA	Yes	Yes	No[3]

1. Normalized with the Min-Max method.

2. BOD requires additive aggregation, similar arguments apply to UCM.

3. At least with the multi-criteria methods requiring weights as importance coefficients.

4. With both linear and geometric aggregations weights are trade-offs and not "importance" coefficients.

Weights may also be chosen to reflect the statistical quality of the data. Higher weights could be assigned to statistically reliable data with broad coverage. However, this method could be biased towards the readily available indicators, penalizing the information that is statistically more problematic to identify and measure.

When using equal weights, it may happen that – by combining indicators with a high degree of correlation – an element of double counting be introduced into the composite: if two collinear indicators are included in the composite with a weight of w_1 and w_2, the unique dimension that the two indicators measure would have weight (w_1+w_2) in the composite. The answer has often been to test indicators for statistical correlation and to choose only indicators which exhibit a low degree of correlation, or to adjust weights correspondingly, e.g. giving less weight to correlated indicators. However, when two correlated indicators convey strong messages they can both be considered in the aggregation. Let us give an example. One may be willing to buy a sport car; therefore both speed and design are suitable indicators for the decision; and they are normally highly correlated.

The true symptom of double counting occurs when two indicators are redundant, i.e. they give the same information. For example, in the CI of e-Business Readiness, the indicators "Percentage of firms using Internet" and "Percentage of enterprises that have a website" display a correlation of 0.88 in 2003. If it is believed that the two indicators are redundant, as they are measuring the same aspect of investment in information and communication technologies, then their weights should be reduced.

The existing literature offers a quite rich menu of alternative weighting methods, all having pros and cons. Factor analysis (FA) could be used to group individual indicators according to their degree of correlation. The benefit-of-the-doubt (Melyn and Moesen, 1991, Cherchye et al., 2004) can be seen as a particular case of *data envelopment analysis* (DEA). It calculates the composite indicator for a given country by using the *best* set of weights, which maximises the composite indicator for that country with respect to the best performing country using the same set of weights. The same procedure is followed for each country. Weights are therefore country-dependent. In general, even using the best combination of weights for a given country, other countries may show better performance. The optimization process could lead to many zero weights if no restrictions on the weights were imposed. In such case many countries would have the value of composite equal to one. Bounding restrictions on weights are hence necessary for this method to be of practical use. For the BOD method the lower bound is usually set to 5% and the upper bound to 10%.

The weights can be assigned by participatory methods that incorporate various stakeholders – experts, citizens and politicians. An example is the *budget allocation*, in which experts are given a “budget” of *N* points, to be distributed over a number of individual indicators, “paying” more for those indicators whose importance they want to stress (Jesinghaus, 1997). The budget allocation is optimal for a maximum of 10-12 indicators. If too many indicators are involved, this method can induce cognitive stress in the experts who are asked to allocate the budget. Public opinion polls have been extensively used over the years as they are easy and inexpensive to carry out (Parker, 1991).

Indicators are aggregated in a number of ways. The linear aggregation method is the most commonly used, providing full compensability among the underlying indicators. Geometric aggregations are better suited if the builder wants some degree of non-compensability. In both linear and geometric aggregations, weights express trade-offs between indicators. A deficit in one dimension can thus be offset (compensated) by a surplus in another. This implies an inconsistency between how weights are conceived (usually measuring the importance of the associated variable) and their actual meaning when geometric or linear aggregations are used. In linear aggregation, the compensability is constant, while with geometric aggregations compensability is lower for the indicators with low values. In terms of policy, if compensability is admitted (as in the case of pure economic indicators), a country with low scores on one indicator will need a much higher score on the others to improve its situation when geometric aggregation is used. Thus, in benchmarking exercises, countries with low scores prefer a linear rather than a geometric aggregation. On the other hand, under geometric aggregation, the marginal utility from an increase in low absolute score would be much higher than in a high absolute score. Consequently, a country would have a greater incentive to address those sectors/activities/alternatives with low scores if the aggregation were geometric rather than linear, as this would give it a better chance of improving its position in the ranking (Munda and Nardo, 2005).

The multi-criteria aggregation procedure tries to resolve the conflict arising in country comparisons as some indicators are in favour of one country while other indicators are in favour of another. This conflict can be treated in the light of a non-compensatory logic and taking into account the absence of preference independence within a discrete multi-criteria approach (Munda, 1995). The approach employs a mathematical formulation (Condorcet's ranking procedure) to rank in a complete pre-order (i.e. without any incomparability relation) all the countries from the best to the worst one after a pair-wise comparison of countries across the whole set of the available indicators. We offer here a 'hand waiving' description of the method. Imagine we have just three countries, A, B and C, and we want to compare them with one another. We build to this effect an 'outscoring matrix' whose entries e_{ij} tell us how much country 'i' does better than country 'j'. The entry e_{ij} is in fact the sum of all weights of all indicators for which country 'i' does better than country 'j'. Likewise, e_{ji} will be the sum of all weights for which the reverse is true. If the two countries do equally well on one variable, its weight is split between e_{ij} and e_{ji}. As a result $e_{ij} + e_{ji} = 1$ if weights have been correctly normalised. We now write down all permutations of county order (ABC,ACB,BAC,BCA,CAB,CBA) and compute for each of them the ordered sum of the scores, e.g. for ABC we compute $Y=e_{AB}+e_{AC}+e_{BC}$. We do this for all permutations and take as the multi-criteria country ranking the one with the highest total score Y. Note that this ordering is only based on the weights, and on the sign of the difference between countries values for a given indicator, the magnitude of the difference being ignored. Hence, to exemplify, a country that does marginally better on many indicators comes out better than a country that does much better on a few ones. In principle the opposite is true for the BOD approach. Note that the multi-criteria method is scale-free and no normalisation is required.

To ensure that weights remain a measure of importance, other aggregation methods should be used, in particular methods that do not allow compensability. Moreover, if different goals are equally legitimate and important, a non-compensatory logic might be necessary. This is usually the case when highly different dimensions are aggregated in the composite, as in the case of environmental problems that include physical, social and economic data. If the analyst decides that an increase in economic performance cannot compensate for a loss in social cohesion or a worsening in environmental sustainability, then neither the linear nor the geometric aggregation is suitable. A non-compensatory multi-criteria approach (MCA) could assure non-compensability by finding a compromise between two or more legitimate goals.

In its basic form this approach does not reward outliers, as it retains only ordinal information, *i.e.* those countries having a greater advantage (disadvantage) in individual indicators. This method, however, could be computationally costly when the number of countries is high (Munda an Nardo, 2009).

With regard to the time element, keeping weights unchanged across time might be justified if the researcher is willing to analyse the evolution of a certain number of variables, as in the case of the evolution of the EC Internal Market Index from 1992 to 2002. Weights often do not change in time. If, instead, the objective of the analysis is that of defining best practice or of setting priorities, then weights should necessarily change over time.

The absence of an “objective” way to determine weights and aggregation methods does not necessarily lead to rejection of the validity of composite indicators, as long as the entire process is transparent.

By the end of Step 6 the developer should have:

- Selected the appropriate weighting and aggregation procedure(s) with reference to the theoretical framework.
- Considered the possibility of using alternative methods (multi-modelling principle).
- Discussed whether correlation issues among indicators should be accounted for
- Discussed whether compensability among indicators should be allowed
- Documented and explained the weighting and aggregation procedures selected.

2.7 STEP 7: Robustness and sensitivity

Sensitivity analysis can be used to assess the robustness of composite indicators

Several choices and judgments have to be made when constructing composite indicators, *e.g.* on the selection of indicators, data normalization, imputation, choice of weights and aggregation methods, *etc*. The robustness of the composite indicator scores and of the underlying policy messages may thus be contested. A combination of uncertainty and sensitivity analysis can help gauge the robustness of the composite indicator and improve transparency.

Uncertainty analysis focuses on how uncertainty in some input variables propagates through the structure of the composite indicator and affects the composite indicator scores. Sensitivity analysis assesses the contribution of the individual source of uncertainty to the variance of the composite scores. While uncertainty analysis is used more often than sensitivity analysis and is almost always treated separately, the iterative use of uncertainty and sensitivity analysis during the development of a composite indicator could improve its robustness (Saisana *et al*., 2005a; Tarantola *et al.*, 2000; Gall, 2007). Ideally, all potential sources of uncertainty should be addressed: selection of individual indicators, data quality, normalization, weighting, aggregation method, *etc*. The approach taken to assess uncertainties could include the following steps:

1. Inclusion and exclusion of individual indicators.
2. Assessing uncertainty in the input data.
3. Using alternative editing schemes, *e.g.* single or multiple imputation.
4. Using alternative data normalisation schemes, such as min-max, standardisation, rankings.
5. Using different weighting schemes, *e.g.* methods from the participatory family (budget allocation, analytic hierarchy process) and endogenous weighting (associated to the benefit of the doubt).
6. Using different aggregation systems, *e.g.* linear, geometric mean of un-scaled variables, multi-criteria and benefit of the doubt.
7. Using different plausible values for the weights.

The acknowledgement of the uncertainties in the development of a composite indicator is made in very few studies. In the Human Development Index the results of the robustness analysis are generally reported as country rankings with their related uncertainty bounds, which are due to the uncertainties at play. This makes it possible to communicate to the user the plausible range of the composite indicator values for each country. The sensitivity analysis results are generally shown in terms of a sensitivity measure for each source of uncertainty. These sensitivity measures represent how much the uncertainty in the composite indicator for a country would be reduced if that particular input source of uncertainty could be removed. The results of a sensitivity analysis are often also shown as scatter plots with the values of the composite indicator for a country on the vertical axis and each input source of uncertainty on the horizontal axis. Scatter plots help revealing patterns in the input-output relationships.

By the end of Step 7 the developer should have:

- Identified the sources of uncertainty in the development of the composite indicator.
- Assessed the impact of the uncertainties/assumptions on the final result.
- Conducted sensitivity analysis of the inference, *e.g.* to show what sources of uncertainty are more influential in determining the relative ranking of two entities.
- Documented and explained the sensitivity analyses and the results.

2.8. STEP 8: Back to the details

De-constructing composite indicators can help extending the analysis

Composite indicators provide a starting point for analysis. They provide the big picture, and this is good for the media to provide the general message; but for policy making we cannot rely on single crisp figures. Policy makers need to identify the weak elements on which action should be made. So we need to go back to the details and analyze the underlying components. Tools like Bayesian networks could help to further illuminate the relationship between the composite and its components.

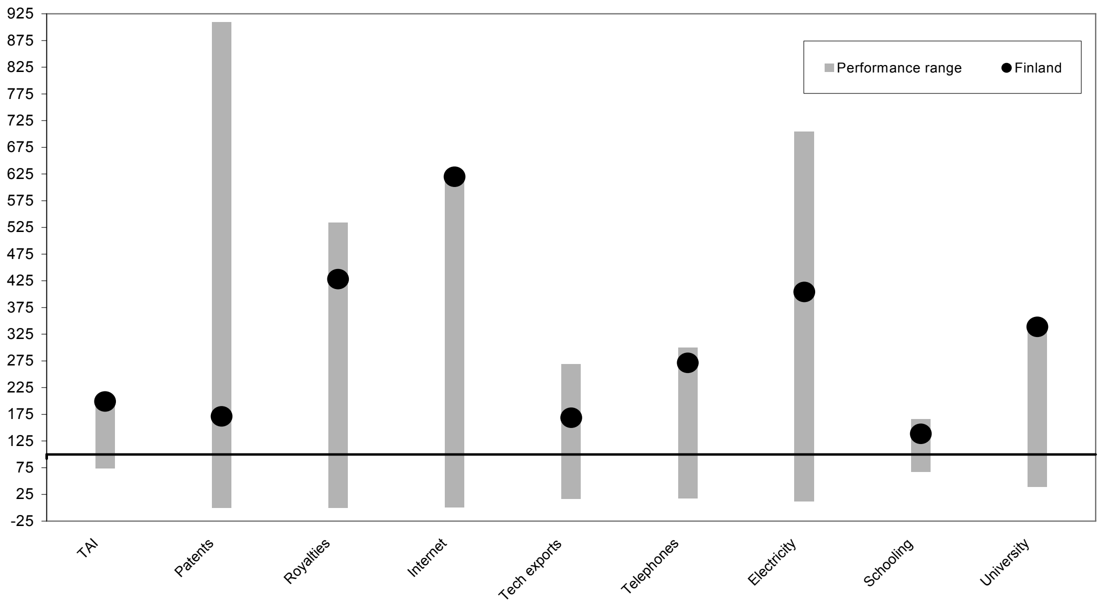
For instance, to profile national innovation performance**,** individual indicators can be used to show strengths and weaknesses of each country. Individual indicators and country profiles can be presented in various ways: (i) leaders and laggards, (ii) spider diagrams and (iii) traffic light presentations.

In the first example (Figure 1), the performance of each indicator can be compared to the leader, the laggard and the average performance. Finland’s top ranking is primarily based on having the highest values for the indicators relating to the *Internet* and *university*, while the country’s only weakness relates to the *patents* indicator.

Another way of illustrating country performance is to use spider diagrams or radar charts (Figure 2). Here Finland is compared to the three best countries on each indicator and to one other country, here the United States.

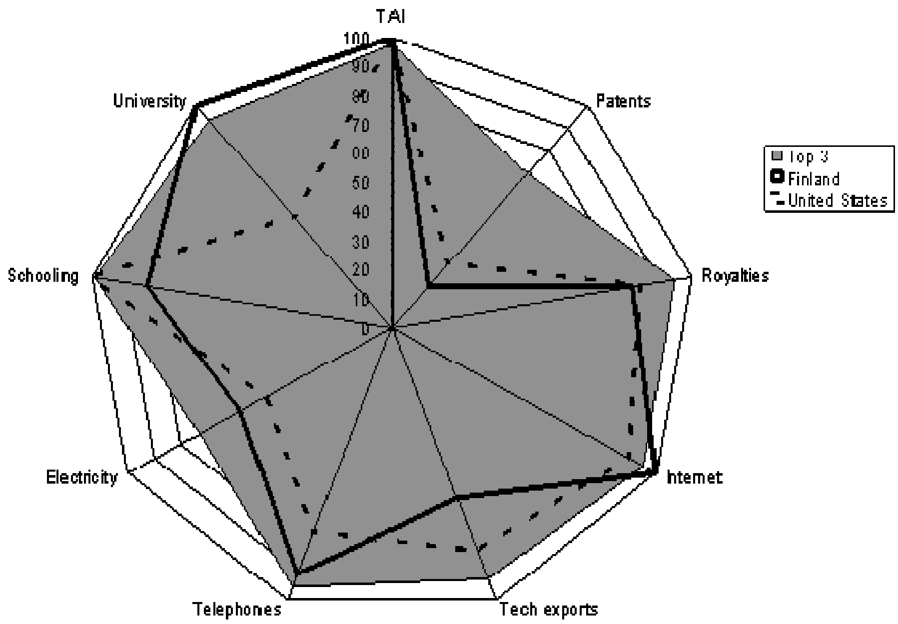
Finally, one can use a traffic light approach, where each indicator takes the color green, yellow or red, according to the relative performance of the country. This approach is useful when many indicators are used in the composite. For example, Figure 3 shows that Finland has only one indicator in red (*patents*) and one in yellow (*electricity*). Japan has three in orange and one in yellow.

Figure 1. Example of leader/laggard decomposition presentation



Note: Technology Achievement Index (TAI). Finland (the dot) is used as an example. The figure is based on standardised indicators using distance to the mean. The grey area shows the range of values for that particular indicator across countries. The average of all countries is represented by the horizontal line at 100.

Figure 2. Example of spider diagram decomposition presentation



Note: Technology Achievement Index (TAI). Finland is compared to the top three TAI performers and to the United States. The best performing country for each indicator takes the value 100, and the worst, 0.

Figure 3. Example of traffic light decomposition presentation

	Index value	Well below average (under 20)	Below average (20-40)	Average (40-60)	Above average (60-80)	Well above average (over 80)
Finland						
TAI	100					
Patents	19	X				
Royalties	80					X
Internet	100					
Tech exports	63				X	
Telephones	90					X
Electricity	57			X		
Schooling	82					X
University	100					
Japan						
TAI	93					X
Patents	100					
Royalties	41			X		
Internet	24		X			
Tech exports	100					
Telephones	76				X	
Electricity	30		X			
Schooling	77				X	
University	36		X			

Note: Technology Achievement Index (TAI). There are several ways to assign colours. In the chosen format five colours are used but this can easily be reduced to three. For example, green might be given to all indicators with values in the first tercile (above or equal the 67% percentile), yellow in the second tercile (above 33% but below 67%), *etc*.

By the end of Step 8 the developer should have:

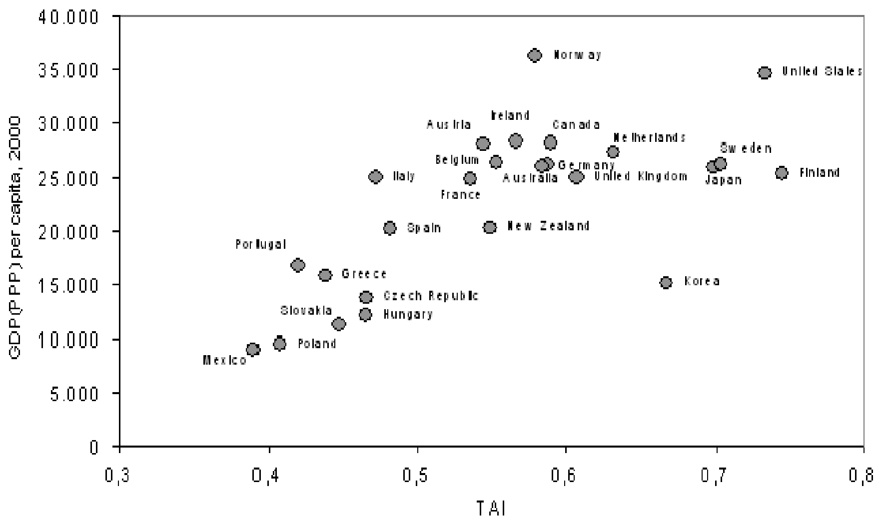
- Decomposed the composite indicator into its individual parts and tested for correlation and causality (if possible) .
- Profiled country performance at the indicator level to reveal what is driving the composite indicator results, and in particular whether the composite indicator is overly dominated by a small number of indicators.
- Documented and explained the relative importance of the sub-components of the composite indicator.

2.9. STEP 9: Links to other variables

Composite indicators can be linked to other variables and measures

Composite indicators often measure concepts that are linked to well-known and measurable phenomena, e.g. productivity growth, entry of new firms. These links can be used to test the explanatory power of a composite. Simple cross-plots are often the best way to illustrate such links. An indicator measuring the environment for business start-ups, for example, could be linked to entry rates of new firms, where good performance on the composite indicator of business environment would be expected to yield higher entry rates. For example, the Technology Achievement Index helps assessing the position of a country relative to others concerning technology achievements. Higher technology achievement should lead to higher wealth, that is, countries with a high TAI would be expected to have high GDP per capita. Correlating TAI with GDP per capita shows this link (Figure 4). Most countries are close to the trend line. Only Norway and Korea are clear outliers. Norway is an outlier due to revenues from oil reserves, while Korea has long prioritized technology development as an industrial strategy to catch up with high-income countries.

Figure 4. Link between TAI and GDP per capita, 2000



Note: The correlation is significantly different from zero at the 1% level and r ² between GDP (PPP, $) and TAI (unitless) equals 0.47. Only OECD countries are included in the correlation, as correlation with very heterogeneous groups tends to be misleading.

One remark is worthwhile at this point. Correlation analysis should not be mistaken with causality analysis. Correlation simply indicates that the variation in the two data sets is similar. A change in the indicator does not necessarily lead to a change in the composite indicator and vice versa. Countries with high GDP might invest more in technology or more technology might lead to higher GDP. The causality remains unclear in the correlation analysis. More detailed econometric analyses can be used to determine causality, *e.g.* the Granger causality test. However, such tests require time series for all variables, which are often not available.

Note also that composite indicators often include some of the indicators with which they are being correlated, leading to double counting. For example, most composite indicators of sustainable development include some measure of GDP as a sub-component. In such cases, the results should be taken with care.

By the end of Step 9 the developer should have:

- Correlated the composite indicator with related measurable phenomena,
- Tested the links with variations of the composite indicator as determined through sensitivity analysis.
- Developed data-driven narratives on the results
- Documented and explained the correlations and the results.

2.10. STEP 10: Presentation and dissemination

A well-designed graph can speak louder than words

The issue of presenting results from composite indicators is not trivial. Composite indicators must be able to communicate a story to decision-makers and other end-users quickly and accurately. Tables, albeit providing the complete information, can sometimes obscure sensitive issues immediately visible with a graphical representation. Therefore we need to decide, in each situation, whether to include a table, a figure, or both. The following examples show three situations where indicator information is communicated graphically. There are plenty of other possibilities. In all situations graphics need to be designed carefully for clarity and aesthetics. In all situations we need to have words, numbers and graphics working together (Trufte, 2001).

A tabular format is the simplest presentation, in which the composite indicator is presented for each country as a table of values. Usually countries are displayed in descending rank order. Rankings can be used to track changes in country performance over time as, for example, the Growth Competitiveness Index, which shows the rankings of countries for two consecutive years (Figure 5). While tables are a comprehensive approach to displaying results, they may be too detailed and not visually appealing. However, they can be adapted to show targeted information for sets of countries grouped by geographic location, GDP, *etc*.

Composite indicators can be expressed via a simple bar chart (Figure 6). The countries are on the vertical axis and the values of the composite on the horizontal one. The top bar indicates the average performance of all countries and enables the reader to identify how a country is performing *vis-à-vis* the average. The underlying individual indicators can also be displayed on a bar chart. The use of colors can make the graph more visually appealing and highlight the countries performing well or not so well, growing or not growing, *etc*.[i]. The top bar can be thought of as a target to be reached by countries.

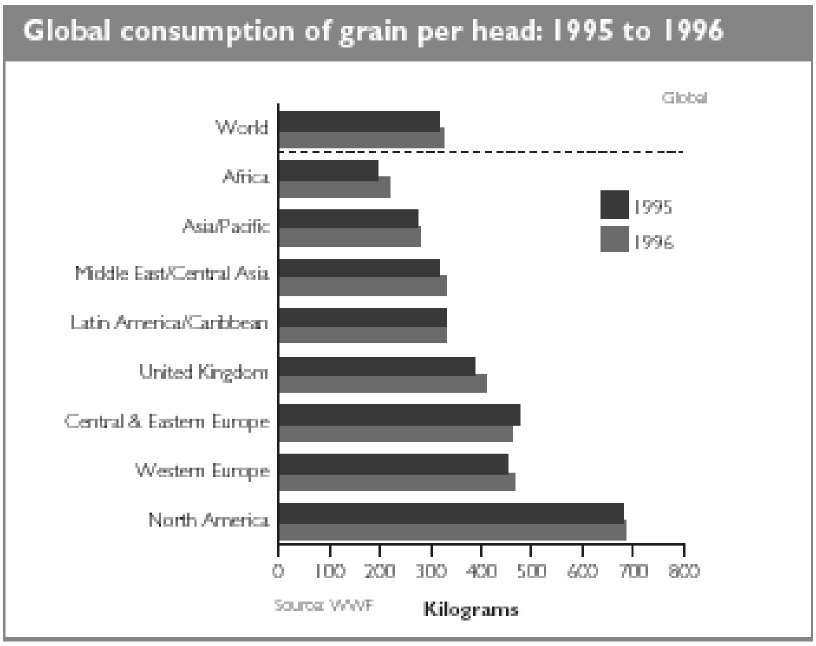
Figure 5. Example of tabular presentation of composite indicator

GROWTH COMPETITIVENESS INDEX RANKINGS

Country	Growth Competitiveness ranking 2003	Growth Competitiveness ranking 2003 among GCR 2002 countries	Growth Competitiveness ranking 2002*
Finland	1	1	1
United States	2	2	2
Sweden	3	3	3
Denmark	4	4	4
Taiwan	5	5	6
Singapore	6	6	7
Switzerland	7	7	5
Iceland	8	8	12
Norway	9	9	8
Australia	10	10	10
Japan	11	11	16
Netherlands	12	12	13
Germany	13	13	14
New Zealand	14	14	15
United Kingdom	15	15	11
Canada	16	16	9
Austria	17	17	18
Korea	18	18	25
Malta	19	—	—
Israel	20	19	17
Luxembourg	21	—	—
Estonia	22	20	27

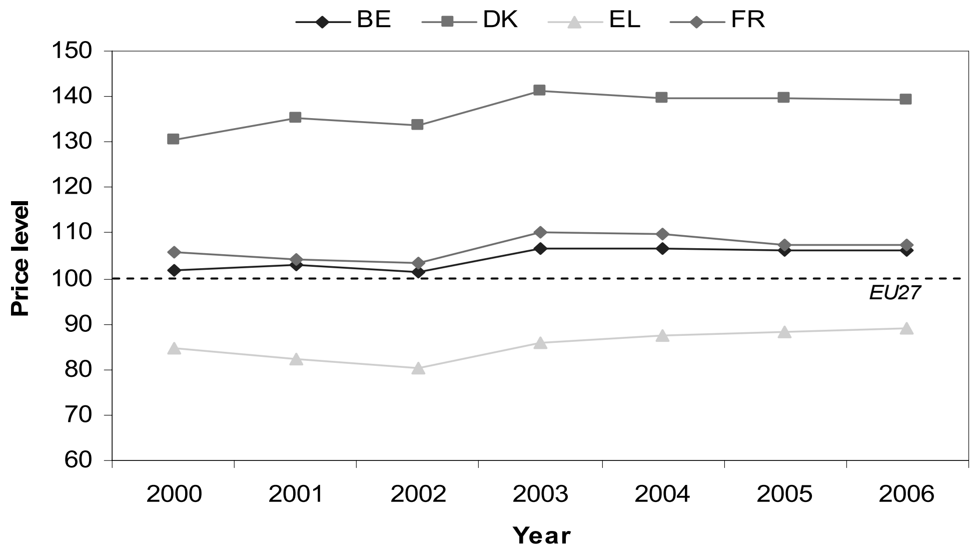
Source: WEF, 2004 www.weforum.org/gcr

Figure 6. Example of bar chart presentation of composite indicator



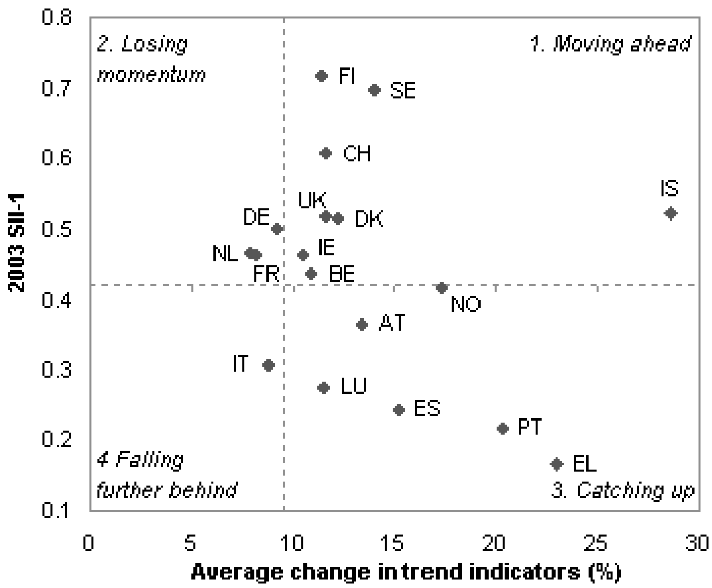
Source: (U.K Government, 2004) http://www.sustainable-development.gov.uk/publications/index.htm

Figure 7. Example of line chart presentation of composite indicator



Note: EU price level index. Comparative price levels of final consumption by private households including indirect taxes (EU-27=100). JRC elaboration, data source: Eurostat, 2007. http:// ec.europa.eu/eurostat

Figure 8. Example of trend diagram of composite indicator



Note: 2003 SII-1 stands for 2003 EU Summary Innovation Index Source: The European Innovation Scoreboard 2003, at ftp://ftp.cordis.europa.eu/pub/focus/docs/innovation_scoreboard_2003_en.pdf For the definition of SII-1 see ibid., p. 9.

Line charts can be used to illustrate the changes of a composite (or its dimensions/components) across time. The values for different countries (or different indicators) are shown by different colors and/or symbols. The indicators can be displayed using, for example, a) absolute levels, b) absolute growth rates, *e.g.* in percentage points with respect to the previous year or a number of past years, c) indexed levels and d) indexed growth rates. When indexed, the values of the indicator are linearly transformed so that their indexed value for a given year is 100 (or another integer). The price level index shows values such that EU27=100 for each year, with more expensive countries having values greater than 100 and less expensive countries below 100 (Figure 7).

Trends in country performance as revealed through a composite indicator can be presented through trend diagrams. When a composite indicator is available for a set of countries for at least two different time points, changes or growth rates can be depicted. The EU Summary Innovation Index is used to track relative performance of European countries on innovation indicators (8). Country trends are reported on the X-axis and levels are given on the Y-axis (although levels in the X-axis and % changes in the Y-axis constitute the usual practice). In this picture, the horizontal dashed line gives the EU average value and the vertical dashed line shows the EU trend. The two lines divide the area into four quadrants. Countries in the upper quadrant are "moving ahead", because both their value and their trend are above the EU average. Countries in the bottom left quadrant are "falling further behind" because they are below the EU average for both variables.

By the end of Step 10 the developer should have:

- Identified a coherent set of presentational tools for the target audience.
- Selected the visualisation technique which communicates the most information.
- Visualised the results of the composite indicator in a clear and accurate manner.

3. The case study: the Active Citizenship Composite Indicator

In the previous section of this paper we have introduced the basic ten guidelines for the construction of a composite indicator. In this section we present an exhaustive case study: the active citizenship composite indicator. This result is part of the "Active Citizenship for Democracy" research project, a project coordinated by the Centre for Research on Education and Lifelong Learning (CRELL), in cooperation with the Directorate General Education and Culture of the European Commission and the Council of Europe. The Active Citizenship Composite Indicator was developed by Hoskins et al., 2006, and the revised version was published in Hoskins and Mascherini, 2009. In what follows, the implementation and the practical application of the ten steps for the construction of the composite indicator are fully documented.

3.1. Step 1: The theoretical framework.

The current European climate has put social cohesion at the heart of the European policy agenda. Active Citizenship is an essential element of this agenda, putting the spotlight on values, representative democracy and civil society. The question is how a concept such as active citizenship can be measured.

In order to give an operational definition of Active Citizenship and to guide the process of the construction of a composite indicator to measure this phenomenon, the "Active Citizenship for Democracy" network was created and coordinated by the Centre for Research on Lifelong Learning (CRELL) of the European Commission. CRELL was created in collaboration between the European Commission's Directorate General for Education and Culture and the Directorate General Joint Research Centre in order to support the monitoring of the Lisbon process in the field of education.

The network was created in cooperation with the Council of Europe's Directorate of Education and it included key experts from across Europe in the fields of social and political science and education. As the concept of Active Citizenship is multidisciplinary, the network of experts included researchers with many different backgrounds.

Building on the foundations of Marshall (1950) in terms of rights and obligations of citizenship and Verba and Nie (1972) in terms of participatory and influential action, the "Active Citizenship for Democracy" network has produced the following definition of Active Citizenship:

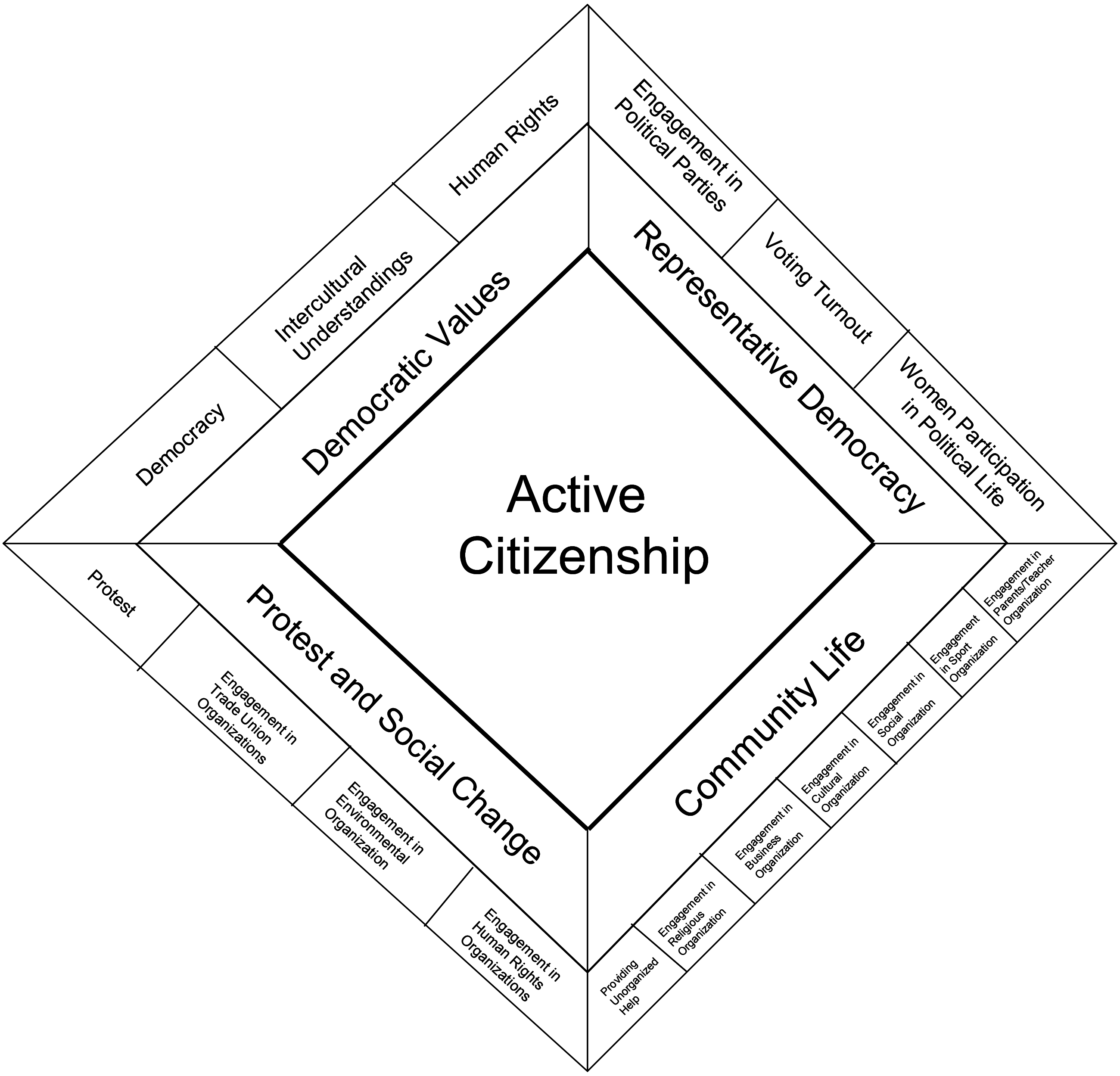
> *Participation in civil society, community and/or political life, characterised by mutual respect and non-violence and in accordance with human rights and democracy.* (Hoskins, 2006[1])

[1] Developed by the CRELL research network "Active Citizenship for Democracy" as part of this project

Active citizenship is partially overlapping with the concept of social values concentrating its interest mostly at the meso and micro-level. Thus, active citizenship is understood in the very broadest sense of the word “participation” and is not restricted to the political dimension. It ranges from cultural and political to environmental activities, on local, regional, national, European and international levels. It includes new and less conventional forms of active citizenship, such as one-off issue politics and responsible consumption, as well as the more traditional forms of voting and membership in parties and NGOs. The limits of active citizenship are set by ethical boundaries. People’s activities should support the community and should not contravene principles of human rights and the rule of law. Participation in extremist groups that promote intolerance and violence should therefore not be included in this definition of active citizenship. As can be seen within this definition, active citizenship incorporates a wide spread of participatory activities containing political action, participatory democracy, civil society and community support. However, and in our view correctly, action alone is not considered active citizenship, the examples of Nazi Germany or Communist Europe can show mass participation without necessarily democratic or beneficial consequences. Instead participation is incorporated with democratic values, mutual respect and human rights. Thus what we are attempting to measure is value based-participation. The difference between this concept and social capital is that the emphasis is placed on the societal outcomes of democracy and social cohesion and not on the benefits to the individual from participation. For further details on the conceptual development of active citizenship we address the reader to Hoskins and Mascherini, 2009.

In order to build the composite indicator of active citizenship in a systematic manner it was necessary to operationalise the definition of the concept. Towards this end we identified measurable and distinctive elements in the definition of active citizenship, which we designated “dimensions of active citizenship.” The dimensions are: participation in Political Life (Representative Democracy), Civil Society (Protest and Social Change), Community Life and the Democratic Values needed for active citizenship (recognition of the importance of human rights, democracy and intercultural understanding). The reason of the double names is the lack of unanimous agreement on these names within the network. In the rest of the paper we will refer to these dimensions using indifferently both names. Furthermore, each dimension was divided into a number of sub-dimensions. The sub-dimensions and basic indicators are obviously influenced by current data availability. When forthcoming surveys provide wider data coverage for active citizenship then the sub-dimensions and base indicators could be refined and improved. The structure of the composite indicator is presented in Figure 9.

Figure 9. Example of line chart presentation of composite indicator



Note: Conceptual framework of the active citizenship composite indicator.

In order to give a full description of the structure of the composite indicator, the dimensions of active citizenship are described and presented more in depth in the next paragraphs.

The dimension of participation in Political Life (Representative Democracy) refers to the sphere of the state and conventional representative democracy such as participation in voting, representation of women in the national parliament and regular party work (party membership, volunteering, participating in party activities and donating money). We did not further divide the dimensions of Political Life into sub-dimensions (as in the other cases), due to the fact that different sources of data were drawn upon. The basic indicators used for this dimension are presented in Table 4.

Table 4. List of basic indicators for the dimension of political life

Political Life Dimension
Description
Political parties: membership
Political parties: participation
Political parties: donating money
Political parties: voluntary work
Worked in political party/action group last 12 months
Donated money to political organisation/action group last 12 months
European Parliament - Voting Turnout
National Parliament - Voting Turnout
Women Participation in national parliament

The dimension of participation in Civil Society (Protest and Social Change) refers in this index to political nongovernmental action. Civil Society has been described as "referring to the arena of un-coerced collective action around shared interests, purposes and values' (Centre for Civil Society 2006). This dimension is based on 18 indicators with the sub-dimensions of protest, human rights organisations, environmental organisations and trade union organisations (the political non-governmental organisations chosen reflect data availability). Protest includes activities such as signing a petition, taking part in a demonstration, boycotting products and ethical consumption. The three subdimensions that refer to NGOs are a combination of indicators on membership, participation in activities, volunteering and donating money. In Table 5 the list of basic indicators for the Civil Society (Protest and Social Change) dimension is shown.

The dimension of participation in Community Life refers to activities that are less overtly political and more orientated towards the community - 'community-minded' or 'community-spirited' activities (Table 6). This dimension could also be understood be comprehended by Civil Society but has been distinguished because these activities are more orientated towards community support mechanisms and less towards political action and accountability of governments. This dimension is based on 25 basic indicators and is divided into seven sub-dimensions: un-organised help, religious organisations, business organisations, sport organisations, cultural organisations, social organisations, parent-teacher organisations (the organisations chosen here reflect data availability). Each sub-dimension referring to an organisation then comprises questions of participation, volunteering, membership and donating money. Some refining of the allocation of basic indicators between the Civil Society and Community Life dimensions may be required.

It could be argued that further survey questions would be needed to feed indicators on informal networks, informal volunteering and family networks. However, apart from the case of non-organised help in the community, data for these types of participation in the community does not currently exist.

It is important to acknowledge at this point that certain characteristics of the definition are difficult to model, e.g. verifying that the participation is non-violent and does not contravene human rights and democracy. This limitation is compensated for by the explicit inclusion of a separate dimension on values.

Table 5. List of basic indicators for the dimension of civil society

Civil Society Dimension

Sub-dimensions	Description
Protest	*Working in an organisation or association*
Protest	*Signing a petition*
Protest	*Taking part in lawful demonstrations*
Protest	*Boycotting products*
Protest	*Ethical consumption*
HR Org.	*Human Rights Organisation: membership*
HR Org.	*Human Rights Organisation: participation*
HR Org.	*Human Rights Organisation: donating money*
HR Org.	*Human Rights Organisation: voluntary work*
TU Org.	*Trade Union Org. : membership*
TU Org.	*Trade Union Org. : participation*
TU Org.	*Trade Union Org. : donating money*
TU Org.	*Trade Union Org. : voluntary work*
Env. Org.	*Environmental Org. : membership*
Env. Org.	*Environmental Org. : participation*
Env. Org.	*Environmental Org. : donating money*
Env. Org.	*Environmental Org. : voluntary work*
Protest	*Contacted a politician*

The dimension of Democratic Values is a combination of indicators on democracy and human rights, the foundation for active citizenship practices, and can be found in the definition of active citizenship. We have also added intercultural understanding because, as highlighted earlier in this report, in the context of a culturally diverse Europe with increasing levels of migration, intercultural understanding is one of the key competences of active citizenship. This is supported by the European Commission's Expert Group on Active Citizenship, which placed intercultural competence as the highest priority of all competences for active citizenship. The possibilities for indicators on human rights are quite limited and this sub-dimension will need to be improved with new data from forthcoming surveys. In total, the dimension of Values was based on eleven basic indicators and divided into three sub-dimensions: human rights, intercultural competencies and democracy. The basic indicators for this dimension are presented in Table 7.

Table 6. List of basic indicators for the dimension of community life

Community Dimension

Sub-dimension	Description
Non-Organised Help	*Non-organised help in the community*
Religious Org.	*Religious Org.: membership*
Religious Org.	*Religious Org.: participation*
Religious Org.	*Religious Org.: donating money*
Religious Org.	*Religious Org.: voluntary work*
Business Org.	*Business Org.: membership*
Business Org.	*Business Org.: participation*
Business Org.	*Business Org.: donating money*
Business Org.	*Business Org.: voluntary work*
Sports Org.	*Sport Org.: membership*
Sports Org.	*Sport Org.: participation*
Sports Org.	*Sport Org.: donating money*
Sports Org.	*Sport Org.: voluntary work*
Cultural Org.	*Cultural Org.: membership*
Cultural Org.	*Cultural Org.: participation*
Cultural Org.	*Cultural Org.: donating money*
Cultural Org.	*Cultural Org.: voluntary work*
Social Org.	*Social Org.: membership*
Social Org.	*Social Org.: participation*
Social Org.	*Social Org.: donating money*
Social Org.	*Social Org.: voluntary work*
Teacher Org.	*Teacher Org.: membership*
Teacher Org.	*Teacher Org.: participation*
Teacher Org.	*Teacher Org.: donating money*
Teacher Org.	*Teacher Org.: voluntary work*

Table 7. List of basic indicators for the dimension of values

Values Dimension	
Sub-dimension	Description
Human Rights	*Immigrants should have same rights*
Human Rights	*Law against discrimination in the work place*
Human Rights	*Law against racial hatred*
Intercultural	*Allow immigrants of different race group from majority*
Intercultural	*Cultural life undetermined/enriched by immigrants*
Intercultural	*Immigrants make country worse/better place*
Democracy	*How important for a citizen to vote*
Democracy	*How important for a citizen to obey laws*
Democracy	*How important for a citizen to develop an independent opinion*
Democracy	*How important for a citizen to be active in a voluntary org.*
Democracy	*How important for a citizen to be active in politics*

3.2. Step 2: The selection of variables

In the field of active citizenship availability of data is a serious problem. Not all dimensions are sufficiently covered and multi-annual data are generally not available. With this in mind, the selection of indicators for the composite measure of active citizenship has been based mostly upon one source of data, which helps to maximize the comparability of the indicators. The source of data chosen was the European Social Survey (http://www.europeansocialsurvey.org/) which ran a specific module on citizenship in 2002. The European Social Survey (ESS) aimed to be representative of all residents among the population aged 15 years and above in each participating country. The size and the quality of the sample make the country coverage of Europe in the ESS data reasonably good, with 19 European countries, including 18 EU member states, providing sufficient quality of data.

Overall, the Active Citizenship Composite Indicator presented in this paper is based on a list of 61 basic indicators. This set of 61 indicators has been chosen strictly following suggestions and recommendations of the Active Citizenship fur Democracy network of experts. The process of the selection of the basic indicators has been very difficult due to the different point of view and expertise of our network. Nevertheless, and as stated above, most of these indicators use individual data collected in the European Social Survey of 2002. In addition, voter turnout at national and European elections has also been considered, as well as the proportion of women in national parliaments. In order to complete the dataset, one missing value has been imputed for Norway. The list of the 19 countries included in the analysis is given in table 8 below. The list of the basic indicators can be found in Hoskins and Mascherini 2009.

Table 8: List of countries that have been analysed

List of Countries			
Austria	Netherlands	Finland	Slovenia
Italy	Denmark	Portugal	Greece
Belgium	Norway	France	Ireland
Luxembourg	Spain	Sweden	Hungary
Germany	Poland	United Kingdom	

3.3. Step 3: Imputation of Missing Data

The quality of the set of indicators adopted for the construction of the composite indicator was very high and they presented just one missing value: as Norway is not part of the European Union the voting turn out at the European Parliament Election was not available. In order to maintain this country in the analysis we decided to impute this value by using Multiple Imputation technique based on the EM algorithm. The effect of the inclusion of these imputed values was investigated during the sensitivity analysis step.

3.4. Step 4: Multivariate Analysis

The information inherent in a dataset of sub-indicators that measure the performance of several countries can be studied along two dimensions, i.e. along sub-indicators and along countries, not independently of each other. When we study the dataset along sub-indicators our aim is to assess if the theoretical structure suggested by the theoretical framework is confirmed by data. On other hands, the study of the dataset along countries we will give us information about group of countries presenting similarities in terms of sub-indicators.

Information on sub-indicators.

Factor analysis was used in order to explore whether the theoretical composition of the dimensions and the sub-dimensions was supported by the data. Factor analysis is a statistical technique that identifies underlying factors that explain correlations between the indicators. In this way, we can identify how the different indicators are related to each other within each dimension. A broader introduction to Factor Analysis can be found in Stevens (1986) and Kim, J. and Mueller (1978a, 1978b). The factor analysis was done using the Principal Components extraction method. A varimax rotation was conducted to facilitate the interpretation of the results. By rotating one looks for a so-called 'simple structure' which implies that items have high loadings on as few factors as possible and at the same time factors have many high and many low loadings. Varimax rotation is an orthogonal rotation resulting in independent, uncorrelated factors.

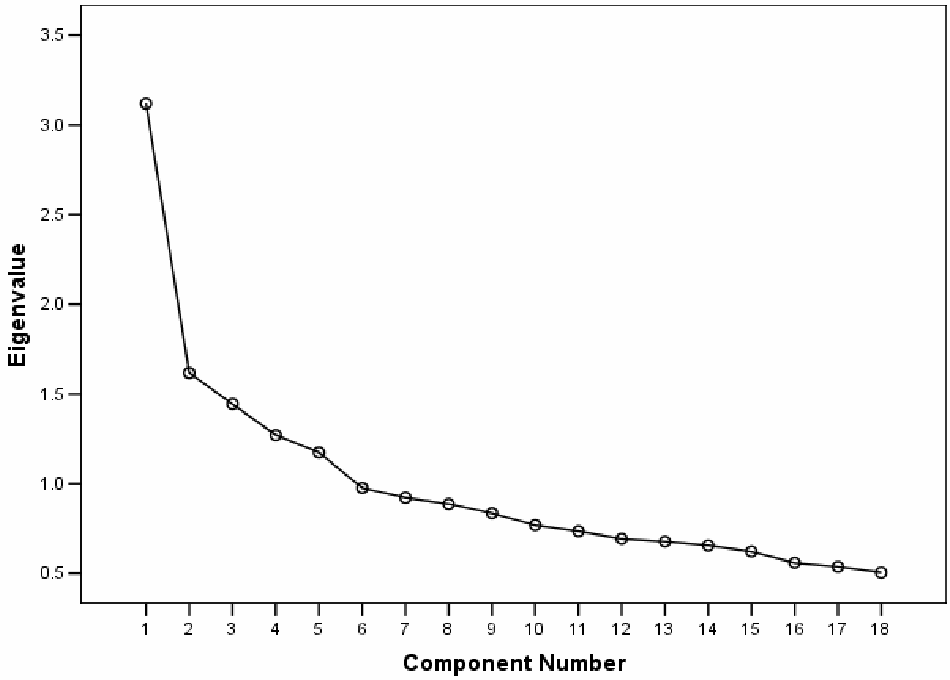
Civil Society (Protest and Social Change)

Eighteen indicators were included for the Civil Society (Protest and Social Change) dimension. The factor analysis shows that five components have eigenvalues greater than 1. These components jointly explain 48% of the variance. In Table 9 the factor loadings for each of the indicators on the components are shown. The first component encompasses indicators referring to protest activities, such as “having signed petitions in the last 12 months” or “boycotted certain products for political/ethical reasons.” The second component refers to people that are members of, participate in, donate money to and do voluntary work for trade unions. The third component groups indicators referring to humanitarian organisations. The fourth component is more difficult to interpret. It has a negative loading for boycotting products for political reasons and positive loadings for membership and donating money to environmental and humanitarian organisations. To some extent, the component refers to people that are involved in civil society in a somewhat passive way. They provide money to certain types of organisations but they do not boycott products or behave actively in other form of participation. The fifth component groups indicators on environmental, peace or animal organisations. Except for the passive participation element, all the other components were hypothesised in the original theoretical structure of the Active Citizenship Composite Indicator. In figure 10, the scree plot of the eigenvalues for the dimension of Civil Society (Protest and Social Change) is presented

Table 9 – Factor loadings for the Civil Society (Protest and Social Change) Dimension

	Component				
	1	2	3	4	5
S1	**0.53**	-0.15	-0.23	0.03	-0.09
S2	**0.65**	-0.02	-0.03	-0.17	-0.02
S3	**0.56**	-0.03	-0.06	0.16	-0.16
S4	**0.63**	0.03	0.04	-0.31	0.03
S5	**0.60**	-0.01	0.00	**-0.42**	0.07
S18	**0.50**	-0.10	-0.11	0.10	-0.09
S6	-0.14	0.06	**0.58**	0.29	-0.06
S7	-0.09	0.01	**0.74**	0.00	0.15
S8	-0.07	0.09	0.33	**0.63**	-0.10
S9	-0.06	0.02	**0.77**	0.00	0.08
S10	-0.11	0.03	0.03	**0.47**	**0.42**
S11	-0.14	0.02	0.08	0.07	**0.73**
S12	0.00	0.03	-0.01	**0.71**	0.26
S13	-0.05	0.02	0.07	0.06	**0.74**
S14	-0.19	**0.55**	-0.02	0.21	-0.14
S15	-0.07	**0.73**	0.04	-0.01	0.04
S16	0.04	**0.66**	0.00	0.05	0.05
S17	-0.05	**0.66**	0.07	-0.08	0.07

Figure 10 – scree plot of eigenvalues for the dimension of Civil Society



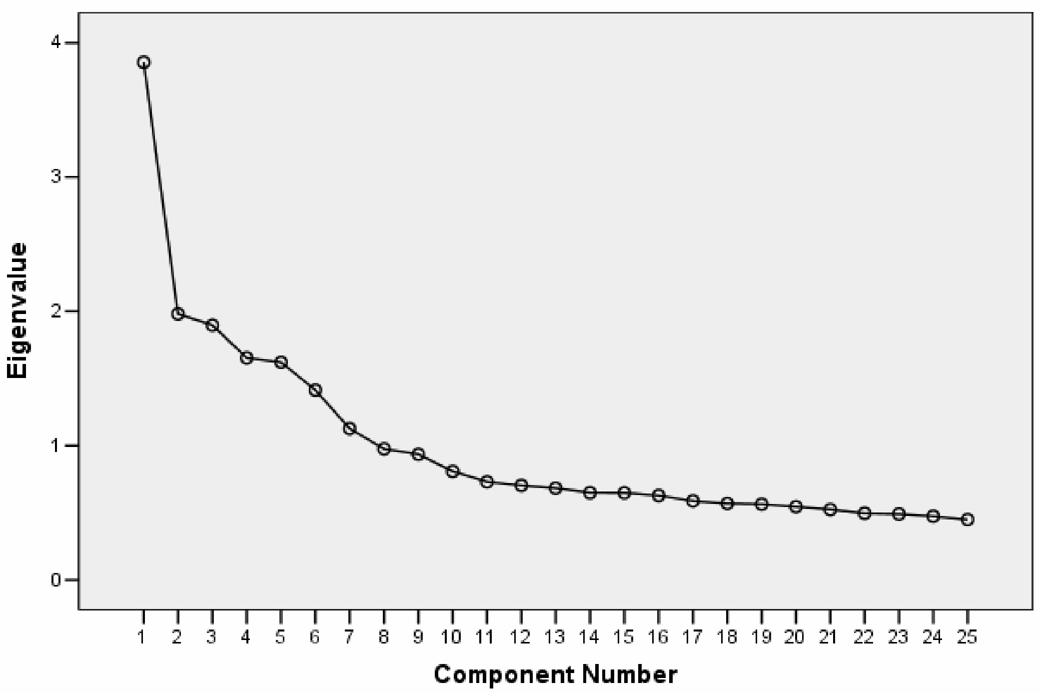
Community Life

The dimension Community Life consisted of 25 indicators referring to membership, participation, voluntary work and having donated money to different types of organisations with an extra indicator for providing help which is not part of the organised voluntary work. The factor analysis shows seven components with eigenvalues greater than 1. The seven components jointly explain 54 percent of the variance. The factor loadings confirm that community-minded action is divided into different subgroups following the applied theoretical structure. The first six components refer to different organised forms of community participation. The Factor Analysis clearly shows that these are distinct modes of community participation since there is no overlap in the components. The seventh component has a negative loading for non-organised support in the community, and positive loadings for the different indicators of membership of a certain organisation. The results show that people who are not members of organisations are those who are more likely to report themselves as helping in a non-organised volunteering context. On table 10 and Figure 11, factor loadings and Eigenvalues scree plot are presented.

Table 10 – Factor loadings for the Community Life dimension

	Component						
	1	2	3	4	5	6	7
S19	-0.10	-0.04	-0.05	-0.08	-0.07	-0.04	**-0.39**
S20	**0.65**	0.01	0.02	0.02	-0.01	0.03	**0.31**
S21	**0.78**	0.06	0.06	0.03	0.04	0.03	0.03
S22	**0.77**	0.06	0.06	0.06	0.08	0.05	-0.07
S23	**0.73**	0.08	0.08	0.01	0.06	0.02	0.00
S24	0.00	0.00	0.04	**0.66**	0.01	0.02	**0.41**
S25	0.03	0.06	0.08	**0.75**	0.06	0.04	0.12
S26	0.09	0.08	0.16	**0.63**	0.07	0.12	-0.26
S27	0.01	0.06	0.11	**0.73**	0.07	0.04	0.01
S28	0.02	0.05	**0.66**	0.04	0.01	0.03	**0.43**
S29	0.06	0.09	**0.71**	0.13	0.08	0.05	0.13
S30	0.11	0.09	**0.67**	0.11	0.08	0.10	-0.23
S31	0.05	0.10	**0.75**	0.10	0.06	0.03	0.00
S32	0.02	-0.01	0.02	0.04	0.05	**0.65**	0.29
S33	0.02	0.04	0.06	0.06	0.07	**0.75**	0.05
S34	0.05	0.01	0.05	0.05	0.04	**0.68**	-0.12
S35	0.02	0.04	0.04	0.03	0.05	**0.68**	-0.05
S36	0.01	**0.66**	0.05	-0.01	-0.02	0.00	0.38
S37	0.04	**0.76**	0.09	0.07	0.04	0.02	0.10
S38	0.10	**0.67**	0.06	0.07	0.06	0.05	-0.21
S39	0.07	**0.74**	0.11	0.07	0.07	0.03	-0.03
S40	0.01	-0.01	0.04	0.02	**0.65**	0.06	**0.34**
S41	0.04	0.04	0.06	0.06	**0.74**	0.06	0.10
S42	0.09	0.05	0.04	0.04	**0.63**	0.06	-0.19
S43	0.04	0.06	0.07	0.07	**0.75**	0.04	0.00

Figure 11 – scree plot of eigenvalues for the dimension of Community Life



Democratic Values

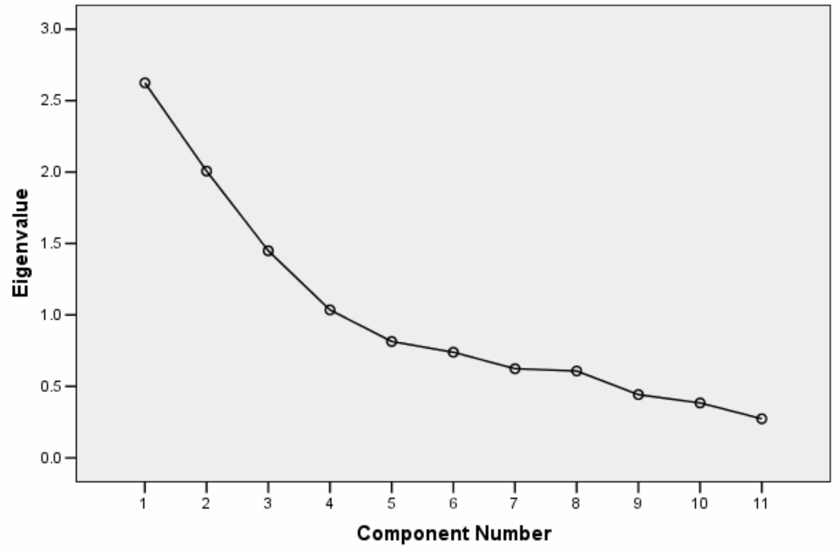
The dimension of Values, in relation to democracy and human rights, was composed of 12 indicators. Within this dimension two analyses were carried out. The first analysis identified four components with eigenvalues greater than 1. Because the eigenvalue of the fourth component is very close to 1, and because a solution with three components might be more in line with the theoretical assumption about the sub-dimensions within the Values dimension, a second FA was carried out considering only three components. In both FA analyses the results were similar. The only difference is that in the first FA with four components the sub-dimension democracy is split up into two groups. Table 11 shows the loadings for a solution with three components. The first component captures positive attitudes towards immigrants, confirming the subdimension of intercultural understanding. The second refers to attitudes towards democracy. The third captures human rights. These three components confirm the theoretical structure except in the case of indicator S44 (i.e. that immigrants should be given same rights as everyone else) which shifts from the human rights subdimension to the intercultural understanding sub-dimension. Figure 12 presents the eigenvalues scree plot for the dimension of Democratic Values

Table 11 – Factor loadings for the Democratic Values Dimension

	Component		
	1	2	3
S44*	**-0.54**	-0.09	-0.08
S45	0.15	0.05	**0.91**
S46	0.13	0.05	**0.91**
S47*	**-0.73**	0.05	-0.07
S48	**0.80**	0.04	0.12
S49	**0.82**	0.05	0.04
S50	0.06	**0.71**	0.06
S51	-0.15	**0.59**	0.10
S52	0.07	**0.55**	0.14
S53	0.07	**0.69**	-0.07
S54	0.10	**0.70**	-0.11

* Reverse scale

Figure 12 – scree plot of eigenvalues for the dimension of Democratic Values



Political Life (Representative Democracy)

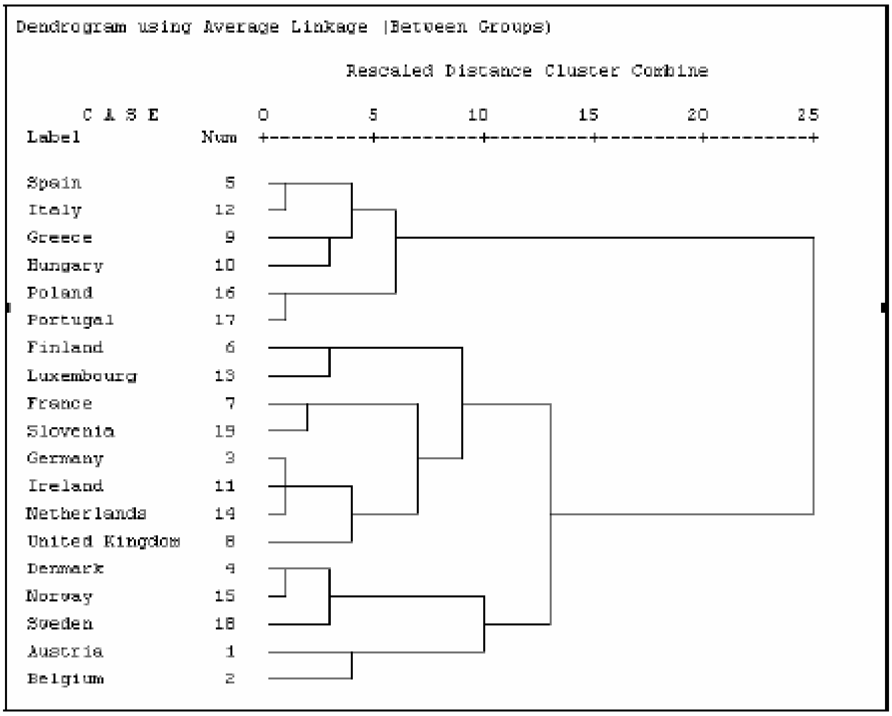
The political life dimension was a combination of nine indicators. Since three of the nine indicators stem from sources other than the ESS1, it was not possible to conduct a factor analysis to confirm the assumption of any structure for this dimension.

In conclusion, it can be said that the different factor analyses within each of the dimensions corroborate the theoretical structure. In other words, the statistical structure of the data corresponds to the theoretical structure.

Information on Countries

This section will investigate what groups can be distinguished among the 19 countries under investigation. For this a technique called cluster analysis is applied. The results of the cluster analysis are presented by means of a dendrogram (see figure 13) in which the clustering of the objects is presented in different steps (there is no ranking in the dendrogram – it shows only which countries are performing in similar ways). The results in the dendrogram clearly show that there are three relatively homogeneous groups. A first group can be seen at the bottom of the dendrogram. This group includes five countries that are regularly found in the group of high-performing countries, namely Sweden, Norway, Denmark, Austria and Belgium. There is a large group of countries which have mid-range scores in the Active Citizenship Composite Indicator. Within this group there is the sub-group of Germany, Ireland, the Netherlands and the UK, a sub-group of France and Slovenia and a sub-group of Finland and Luxembourg. The third group of countries is formed by the southern European countries together with Hungary and Poland.

Figure 13 – Dendrogram of the results of the cluster analysis



The three different clusters of countries presented above give substantial confirmation that the results obtained in the Active Citizenship Composite Indicator are an accurate reflection of the score in the basic indicators as the clusters of countries can be found together in the order of the ranking.

3.5. Step 5: Normalization of Data

At this step variables must be standardized and the weighting scheme for the indicators specified. Due to the fact that the 61 basic indicators have been constructed using different scales, a standardization process is needed before the data for the different indicators can be aggregated. Different standardization techniques are available for this (Nardo et al., 2008). The basic standardization technique that has been applied is the well known z_score approach in which for each basic indicator $x_{m,n}$, the average across countries $\overline{x}_m$ and the standard deviation across countries σ_{x_m} are calculated. The normalization formula is:

$$I_{m,n} = \frac{x_{mn} - \overline{x}_m}{\sigma_{x_m}}.$$

Before the standardization process, the data have then been transformed to ensure that for each indicator a higher score would point to a better performance. This step was clearly necessary to make a meaningful aggregation of the different indicators.

3.6. Step 6: Weighting and Aggregation

Weighting

In the construction of a CI the weighting phase is highly debated in the literature and there is no agreed methodology on how to weight individual indicators. Weights can heavily influence the outcome of a CI and countries ranking. Therefore, weights should ideally be selected according to an underlying and agreed, or at least clearly stated, theoretical framework. Whatever method is used to derive weights, based either on statistical models or on elicitation of expert judgment, no consensus is likely to exist. This is because the weighting implies a "subjective" evaluation, which is particularly delicate in case of complex, interrelated and multidimensional phenomena.

In many CIs all variables are given the same weight when there are no statistical or empirical grounds for choosing a different scheme. Equal weighting (EW) could imply the recognition of an equal status for all sub-indicators. Alternatively, it could be the result of insufficient knowledge of causal relationships, or ignorance about the correct model to apply. In any case, EW does not mean no weighting or a neutral position on the importance of the indicators, because it implies an implicit judgment on the weights being equal. Nevertheless, indicators could also be weighted based on the opinion of experts, who have theoretical knowledge. Theoretical knowledge of Active Citizenship in the development of the weighting scheme would enable the final measurement of the CI to reflect closer the definition and framework of Active Citizenship. It also facilitates the input of the multiplicity of stakeholders' viewpoints into the measurement of Active Citizenship. There are of course fundamental considerations around whose preferences that will be used in the application of those weights and how the preferences of those individuals (or groups of individuals) will be elicited. For instance, where the dimensions of a CI are very technical in nature, the use of "expert" opinion has been advocated. Even if the expert judgment elicitation can not be considered as a substitute for definitive 'objective' scientific research, it can however provide a more systematic representation of the diversity of expert judgments that is typically provided in an experts meeting. In addition, the results of such a study provide a clear indication of the nature and the extent of agreement within the scientific community and also allow conclusions to be drawn about how important the range of expert opinions is to the overall debate, Kahneman et al. (1982) and Seaver (1978).

The members of the Active Citizenship Research Network were unwilling to accept the idea of using weights based on mere statistical techniques, such as equal weighting of factor analysis weights. Nevertheless, some of the members were favourable to being involved in the construction of the CI by utilising their knowledge of the computation of the weights of the basic indicators. In order to permit the elicitation of the experts' judgment, on February 2007 we distributed a questionnaire to 27 leading experts on Active Citizenship. All of the people contacted for participating in the survey had been established as researchers or key experts in the field of the Active Citizenship domain and for this reason they were considered experts. In particular, the participants to the survey belong to four different areas of expertise: sociologists, political scientists, policy makers and educationalists.

The questionnaire was designed following the budget allocation approach, that is a participatory method in which experts are given a ‘‘budget’’ of N points (in our case 100), to be distributed over a number of sub-indicators, paying more for those indicators whose importance they want to stress. (Jesinghaus, 1997). Since the high number of indicators was misleading for an unbiased elicitation of the judgment, Bottomley et al. (2000), the overall questionnaire was composed of six questions where experts were asked for an evaluation of the level of importance of the factors that form a part of the different dimensions. An additional question asked to experts to differentiate the levels of participation in organisations (membership, voluntary work, donating money, participation).For each question each expert was given 100 points and was asked to allocate the budget to the different items according to their importance.

The summary statistics of the level of importance assigned by the experts to the different dimensions of the ACCIs are shown in Table 12. For each expert, the weights of the basic indicators were computed by a linear combination of normalised values of the median of the distribution of the weights assigned to dimensions and sub-dimensions. We chose the median because it is a more robust estimator of central tendency than the mean whose value can be more easily affected by outliers. The medians are then rescaled in order to make the sum of the weights equal to 1. For a detailed description of the computation of the weights and the experts' elicitation process we address the reader to Mascherini and Hoskins (2008).

Aggregation

The structure of the Active Citizenship Composite Indicator is a weighted sum of the indices computed for the four dimensions Di (Political Life, Civil Society, Community, Values):

The structure of the Active Citizenship Composite Indicator is a weighted sum of the indices computed for the four dimensions Di (Political Life, Civil Society, Community, Values):

$$Y_c = \sum_{i=1}^{4} w_i D_{ic} \; ,$$

where $\sum_{i=1}^{4} w_i = 1$ and $0 \le w_i \le 1$ for all i=1,..4, and c=1,..,19, where c indicates the number of countries.

Then, each dimension index, Di, is computed as a linear weighted aggregation of the sub-dimension indices SDij. with weights wj*

$$D_{ic} = \sum_{j=1}^{k_i} w^*{}_j SD_{ijc} \; ,$$

where $\sum_{j=1}^{k} w^*{}_j = 1$ and $0 \le w_j \le 1$ for all j=1,..ki, and again the country index c=1,..,19. The value of ki varies among the different domains Di, and it corresponds to the number of sub-dimensions encompassed by that domain. So, for instance, for the Civil Society domain (i=1), K1 is equal to 4 and for the Community Life Domain (i=2), k2 is equal to 7.

Finally, each sub-dimension index SDij is a linear weighted sum of the sij normalised sub-indicators $I_{h_{i,j}c}$ with weights $w^{\#}_{h_{i,j}}$

$$SD_{ijc} = \sum_{h_{ij}=1}^{s_{ij}} w^{\#}{}_{h_{i,j}} I_{h_{i,j}c} \; .$$

Aggregating the different equations into one gives the general formula for the Active Citizenship Composite Indicator:

$$Y_c = \sum_{i=1}^{4} w_i \sum_{j=1}^{k_i} w_j^* \sum_{h_{ij}=1}^{s_{ij}} w_h^{\#} I_{h_{i,j}c}$$

Having defined the structure, the construction and evaluation of the composite indicator (CI) involve several steps. The first step is the data selection and, if necessary, the imputation of missing data. In the next step the variables must be standardised and the weighting scheme for the indicators specified. Finally, the calculation of the CI and an analysis of its robustness must be performed to improve the transparency of the process.

Table 12 – Results of the Budget Allocation approach

Domain	Item	Expert weight (rescaled median)	Equal weights	Standard deviation	Min	Max
Active Citizenship	Representative democracy	25.81	25	11.28	13	66
	Protest and social change	25.81	25	6.43	11	40
	Community life	22.58	25	7.49	5	35
	Democratic values	25.81	25	7.34	10	40
Engagement	Voluntary work	34.15	25	8.16	10	50
	Participation	34.15	25	10.03	1	50
	Membership	20.32	25	9.46	10	50
	Donating money	11.38	25	9.05	0	33
Representative democracy	Being active in political parties	24.74	33.33	11.64	0	50
	Voting to National/ European Election	39.92	33.33	8.86	22	60
	Women participation	35.34	33.33	9.92	10	50
Protest and social change	Protest	23.07	25	7.87	1	35
	Engagement in human rights organizations	27.58	25	4.83	20	40
	Engagement in trade union organizations	22.70	25	6.14	10	33
	Engagement in environmental organizations	26.64	25	5.10	20	40
Community life	Providing unorganized help	15.86	14.28	7.18	5	32
	Engagement in religious organizations	10.47	14.28	5.06	0	16
	Engagement in business organizations	8.38	14.28	5.11	0	16
	Engagement in sport organizations	10.47	14.28	5.25	1	20
	Engagement in cultural organizations	15.71	14.28	7.59	1	40
	Engagement in social organizations	20.26	14.28	7.08	10	37
	Engagement in teacher/ parents organizations	18.85	14.28	5.14	9	31
Democratic values	Democracy	34.48	33.33	6.78	25	50
	Intercultural understandings	31.04	33.33	4.33	20	33.33
	Human rights	34.48	33.33	5.78	20	50

Results

The ACCI is computed on the basis of the weights elicited by the experts. For each expert, the CI is computed once for all countries. The score assigned to each country corresponds to the median of the distribution of the scores assigned to that country by all the experts (Table 13). Firstly, the overall ranking is introduced and then the rankings for each dimension are presented. Finally, the correlations among the dimensions are shown.

Table 13 – Results of the Active Citizenship Composite Indicator

Rank	Country	Score (median)
1	Sweden	1.017
2	Norway	0.731
3	Denmark	0.600
4	Belgium	0.565
5	Austria	0.436
6	Luxembourg	0.324
7	Netherlands	0.312
8	Germany	0.295
9	Ireland	0.121
10	Finland	0.056
11	United Kingdom	−0.018
12	France	−0.286
13	Spain	−0.352
14	Italy	−0.470
15	Slovenia	−0.474
16	Portugal	−0.565
17	Greece	−0.789
18	Poland	−0.806
19	Hungary	−0.833

Overall, it can be seen that the Nordic countries Sweden, Norway and Denmark score the highest. The exception to this trend is Finland, which for the overall composite and the three dimensions of participatory engagement ranks in the middle of the table. In the domain of Values, however, Finland is ranked 3rd. The group of Scandinavian Countries is followed by Central European Countries: Among them, the highest score is recorded by Belgium, followed by Austria and Netherlands, Luxembourg and Germany. The group of Anglo-Saxon countries plus Finland is ranked from the 9th to the 11th position and they perform much better than France, Mediterranean countries and Slovenia. Finally, in general, it is Eastern Europe and Greece that figure in the lower end of the ranking.

The results among the different dimensions are shown in Table 14. In general, Nordic Countries (especially Sweden) show top performances in all the different dimensions, presenting a valuable consistency in their performances. In contrast, Central European Countries show performances with different profiles; whereas the Netherlands and

Luxembourg have consistent performances in all dimensions considered, Belgium compensates for low scores in the dimension of Values with outstanding performance in Political Life.

Moreover, looking at the individual indicator included in the dimension of Protest and Social Change (Civil Society), the Nordic countries, where NGOs thrive, have high scores, and they are followed by Western European countries. The lower-scoring countries are from Eastern and Southern Europe. The driver of this result is mainly the sub-dimension of protest which is relatively high for all countries considered, whereas the Achilles heel is participation (especially in trades union). The low score of Poland and Hungary is especially driven by a low score for in volunteering working in organisations (6.5% for Poland and 3% for Hungary, compared with the 30% of the top performer) and in participation in human rights organisations (1% for both countries, while the top performer reaches 4.3%). Portugal shows better performance in this latter variable (2%) and Greece is particularly strong in the dimension of protest.

The dimension of Community Life shows a slightly different picture. Here high scores are achieved by Belgium and the UK as well as by the Nordic countries. Participation and membership in sports and cultural activities are the driving force of the result. The low position of Italy is mainly the result of low participation and voluntary work and Spain compensates for its low score in participation and membership with high scores for parent–teacher organisations. For Southern Europe, the variable non-organised help is probably not sufficient to represent the informal networks and family support that characterise this region. In countries like Italy, for example, activities like preserving the food heritage (e.g. the Slowfood movement), or keeping cities lively with evening street activities could be considered relevant. Community participation scores low in Eastern Europe, especially in Poland. Furthermore, in Poland religious activities are more frequent than elsewhere in Europe. The dimension of Democratic Values shows a significantly different pattern from the previous dimensions, with some countries demonstrating quite different behaviour and overall fewer regional distinctions. Poland scores quite well in this index and enters the top five. In contrast to the other dimensions, Portugal also scores well in eighth place. In addition, Finland and Luxembourg join Sweden on the top three. The position of Belgium results from its relatively lower scores in the indicators on values on human rights as only about 2/3 of Belgian respondents said that they would give the same rights to immigrants and about the same number considered important the approval of laws against discrimination in the workplace or against racial hatred. In Sweden the proportions were closer to 90% and 80%, respectively.

Finally, in the dimension of Representative Democracy (Political Life), Austria and Belgium achieve high scores along with the Nordic countries. Austria is ahead of the Nordic countries (in spite of a relatively lower value for women's participation in national parliament), the only occasion in all four dimensions of Active Citizenship that this region does not score the highest. Austria's high score is partly due to the very high number of persons who are involved in political parties. Belgium ranks high in this dimension as a result of its policy of compulsory voting. France and UK perform less well in this dimension than in the previous two indices. Eastern European and some Southern European countries have lower scores. Poland has low voting scores but performs relatively well in donating money to political organisations, whereas Hungary performs well in democratic values and voting (75% in national elections and 38% in European parliament elections) but not in participation in politics. Overall the countries that perform better are not those with the highest voting rates for national or European parliaments but those where participation in politics is higher.

Finally, the correlation ratio for pairs of dimensions of the ACCI is analyzed in Table 15. It is important to note that the correlations are carried out at country level; this means that we are able to discover relationships between country scores and not between the behaviors of individuals.

The ACCI has the highest correlation with the dimension of Protest and Social Change (r & 0.96). High correlation is also found between the dimensions of Community Life (r & 0.91.) and Representative Democracy (r & 0.83.). However, the level of correlation between the overall ACCI and the dimension of Values (r & 0.48), and between Values and the other dimensions of Active Citizenship is not strong. Therefore, whereas the dimensions of Protest and Social Change, Community Life and Representative Democracy seems to move together, showing strong correlation between each other, the dimension of Values seems to display a different and autonomous behavior. This aspect is worthy of more attention in future research.

Table 14 – Rankings of the different dimension of the Active Citizenship Composite Indicator

Rank	Country	Protest and social change	Community life	Democratic values	Representative democracy
1	Sweden	2	2	1	2
2	Norway	1	1	4	7
3	Denmark	3	6	7	3
4	Belgium	4	3	18	1
5	Austria	5	9	9	4
6	Luxembourg	11	10	2	5
7	Netherlands	6	5	11	8
8	Germany	7	7	10	6
9	Ireland	10	8	6	13
10	Finland	12	13	3	9
11	United Kingdom	8	4	13	15
12	France	9	11	16	16
13	Spain	14	14	12	10
14	Italy	15	17	15	11
15	Slovenia	13	12	14	17
16	Portugal	16	15	8	14
17	Greece	18	18	19	12
18	Poland	19	19	5	19
19	Hungary	17	16	17	18

Table 15 – Correlation among the different dimensions

	Active Citizenship (ACCI)	Protest and social change	Community life	Democratic values	Representative democracy
Active Citizenship (ACCI)	1.000	0.959	0.910	0.481	0.833
Protest and social change		1.000	0.926	0.332	0.745
Communities life			1.000	0.284	0.632
Democratic values				1.000	0.197
Representative democracy					1.000

3.7. Step 7: Sensitivity and Robustness Analysis

To investigate the robustness of the ranking based on the composite indicator, the rankings based on several methods of weighting, structures and standardisation methods were compared.

Every aggregate measure or ranking system, including the Active Citizenship Composite Indicator, involves subjective judgments in the selection of indicators, the choice of aggregation model, and the weights applied to the indicators. Because the quality of a ranking system depends on the soundness of its assumptions, good practice requires evaluating confidence in the system and assessing the uncertainties associated with its development process. To ensure the validity of the messages conveyed by this composite indicator, it is important that the sensitivity of the EU country rankings to the structure and aggregation approach be adequately studied.

By acknowledging a variety of methodological assumptions that are intrinsic to policy research, a "sensitivity analysis" can determine whether the main results change substantially when those assumptions are varied over a reasonable range of possibilities (Saisana et al., 2005; Saltelli et al., 2000). Using sensitivity analysis, we can study how variations in rankings derive from different sources of variation in the assumptions. Sensitivity analysis also demonstrates how each model/system depends upon the information that composes it. It is thus closely related to uncertainty analysis, which aims to quantify the overall uncertainty in a country's rank as a result of the uncertainties in the ranking system construction. A combination of uncertainty and sensitivity analyses can help to gauge the robustness of the composite indicator results, to increase its transparency, to identify the countries whose performances improve or deteriorate under certain assumptions, and to help frame the debate around the use of the Framework.

The validity of the Active Citizenship ranking is assessed by evaluating how sensitive it is to the assumptions that have been made about its structure and the aggregation of the 63 individual indicators. The sensitivity analysis is undertaken with respect to three main sources of uncertainty: (1) dimension structure, (2) weighting method - equal weighting, Factor Analysis, or Benefit of the Doubt, and (3) aggregation approach - non-linear/non-compensatory multi-criteria, or an additive aggregation.

For the sensitivity analysis of the Active Citizenship Composite Indicator we analysed 11 scenarios in total, as listed in Table 16. The first eight scenarios employ a linear aggregation, whilst a multi-criterion non-linear/non-compensatory approach is used in the scenarios numbered 9 to 11. The BoD weights can be used exclusively with the linear aggregation and not with the non-linear/non-compensatory aggregation. The dimension structure is preserved in all scenarios, except 1, 2, 7 and 9. Z-scores standardisation is used to normalise the data prior to the additive aggregation in Scenarios 1,3 and 5 and the MinMax normalisation is used in Scenarios 2,4 and 6. No normalisation is needed in the case of either the Benefit of Doubt weighting approach or the non-compensatory multi-criteria (scenarios 4 to 8).

The two normalisation techniques tested are:

Standardisation (or Z-scores):

For each sub-indicator x_{mn}, the average across countries $\bar{x}_{mn}$ and the standard deviation across countries $\sigma_{x_{mn}}$ are calculated. The normalization formula is: $y_{mn} = \frac{x_{mn} - \bar{x}_{mn}}{\sigma_{x_{mn}}}$, so that all the y_{mn} have similar dispersion across countries. This approach converts all indicators 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 sub-indicators.

Min-max scaling:

Each sub-indicator x_{mn} is transformed linearly in $y_{mn} = \frac{x_{mn} - \min_n(x_{mn})}{\max_n(x_{mn}) - \min_n(x_{mn})}$ where $\min_n(x_{mn})$ and $\max_n(x_{mn})$ are the minimum and the maximum value of x_{mn} across all the countries N. In this way, the normalized indicators y_{mn} have values within [0, 1]. This approach increases the impact of indicators with small range of values to the overall composite indicator, which, depending on the case, could be a desirable or undesirable property.

Table 17 presents the overall rankings for all eleven scenarios. One notices that the overall ranking is not sensitive to any of the four major methodological choices made to develop the composite ranking. In the worst cases, the shift in rank is of two positions, mostly due to the aggregation method (non-linear/non-compensatory multi-criteria). This modest sensitivity is observed for Ireland, Luxembourg, Germany, United Kingdom, Poland and Hungary. Norway and Sweden alternatively occupy the top of the ranking. This outcome produces a high degree of confidence that the Active Citizenship Composite Indicator provides a solid framework for assessing relative performance between European countries in a robust way.

For completeness of the analysis, we study whether the relative performance of the countries within each dimension of Active Citizenship is affected by the method employed to aggregate the information. To this end, Table 17 presents the country rankings in each of the four dimensions of Active Citizenship for the proposed ranking and the shifts in rank under three scenarios, namely S2, S6 and S10. Again the rankings in the four dimensions are quite robust to the methods employed to construct/validate the dimensions of Active Citizenship. In most cases, the shift is of one or two positions, with a few exceptions regarding the Civil Society dimension, in which Finland would improve its rank by five positions when using a BoD weighting approach, whilst the Netherlands would lower its rank by five positions under the non-compensatory multi-criteria aggregation.

Table 16: Methodological scenarios for the development of the Active Citizenship Composite Indicator (EW: Equal weights; FA: Factor Analysis; NCMC: Non-Compensatory Multi-criteria)

Scenario	*Dimension Structure*	*Normalisation*	*Weighting*	*Aggregation*
S1	Not Preserved	Standardisation	EW for all indicators	Additive
S2	Not Preserved	MinMax	EW for all indicators	Additive
S3	Preserved	Standardisation	EW for indicators within dimension	Additive
S4	Preserved	MinMax	EW for indicators within dimension	Additive
S5	Preserved	Standardisation	FA weights within dimension, EW for the dimensions	Additive
S6	Preserved	MinMax	FA weights within dimension, EW for the dimensions	Additive
S7	Not Preserved	None	BoD weights for all indicators	Additive
S8	Preserved	None	BoD weights within dimension, EW for the dimensions	Additive
S9	Not Preserved	None	EW for all indicators	NCMC
S10	Preserved	None	EW for indicators within dimension	NCMC
S11	Preserved	None	FA within dimension, EW for the dimensions	NCMC

Table 17: Country rankings in each of the four dimensions of the Active Citizenship Composite Indicator and shifts in rank under three distinct methodological scenarios. Countries are listed in alphabetical order

Active Citizenship		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
1	Sweden	0	−1	0	−1	0	−1	0	0	0	0	0
2	Norway	0	1	0	1	0	1	0	0	0	0	0
3	Denmark	0	0	0	0	−1	−1	0	0	0	0	0
4	Belgium	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
5	Austria	1	1	1	1	2	2	1	1	1	1	1
6	Luxembourg	−4	−4	−1	−2	−2	−3	−3	0	−4	−3	−3
7	Netherlands	1	1	−1	0	0	0	1	−1	0	−1	−1
8	Germany	−1	−1	−1	−1	−1	0	−2	−1	0	1	1
9	Ireland	2	2	3	3	3	3	2	2	3	3	3
10	Finland	−1	−1	−1	−1	−1	−1	−1	0	−2	0	0
11	United Kingdom	3	3	1	1	1	1	3	0	2	0	0
12	France	0	0	0	0	0	0	0	0	1	0	0
13	Spain	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	0
14	Italy	−1	−1	−1	−1	−2	−2	−1	−1	−1	−1	−2
15	Slovenia	2	2	2	2	2	2	2	2	2	2	1
16	Portugal	0	0	0	0	1	1	0	0	0	0	1
17	Greece	0	0	−1	0	−1	−1	−1	0	−1	0	0
18	Poland	0	0	1	0	1	1	1	0	−1	−1	−1
19	Hungary	0	0	0	0	0	0	0	0	2	1	1

3.8. Step 8: link with other indicators:

In order to better understand the phenomenon of active citizenship, the relationship between the Active Citizenship Composite Indicator (ACCI) and other social and economic indicators was explored. A comparison was made with the Corruption Perceptions index (CPI), GDP per capita, the Human Development Index (HDI), the Social Cohesion Index (SCI), the Global Gender Gap Index and the five benchmarks on education and training (plus investment in education) adopted by the Council (Education) in 2003.

The results are presented in Table 18. Overall, the ACCI shows a high correlation with the Corruption Perceptions index, the Human Development Index and GDP per capita. The correlation is slightly lower with the Global Gender Gap Index and evidence is mixed when the benchmarks in education are considered.

Table 18: Correlation between the Active Citizenship Composite Indicator (and its four dimensions) and some indicators in the social and economic domain

	Active Citizenship				
	Civil society domain	Community domain	Values domain	Political Life	Active Citizenship CI
Corruption Perceptions Index	0.862	0.763	0.432	0.604	0.840
Global Gender Gap Index	0.629	0.581	0.589	0.459	0.695
Human development index 2002	0.84	0.71	0.30	0.68	0.79
Social cohesion index	0.63	0.44	0.23	0.44	0.59
Social cohesion index -2	0.77	0.48	0.35	0.49	0.77
GDP per capita (PPP US$ 2002)	0.83	0.75	0.30	0.65	0.79
Indicators in education and training [2]					
Early school leavers	0.40	0.41	0.13	0.29	0.39
Educational attainment	0.27	0.30	0.02	0.17	0.25
Maths and science graduates	0.25	0.22	0.15	-0.06	0.18
Low achievers	0.44	0.38	0.34	0.28	0.44
Lifelong learning	0.68	0.60	0.52	0.35	0.66
Investment in human resources	0.56	0.40	0.34	0.27	0.49

[2] The variables considered are the following: early school leavers (percentage of the population aged 18-24 with at most lower secondary education and not in further education or training; educational attainment (percentage of population aged 20 to 24 having completed at least upper secondary education); maths and science graduates (tertiary graduates in science and technology per 1000 of population aged 20-29); low achievers (% of pupils at level 1 or below in the PISA literacy scale); lifelong learning (percentage of the adult population aged 25 to 64 participating in education and training); and investment in human resources (public expenditure on education as a percentage of GDP). For further details see the web site http://ec.europa.eu/education/policies/2010/news_en.html

Transparency International Corruption Perceptions Index[3]

The Transparency International Corruption Perceptions Index assesses 163 countries in terms of the degree to which corruption is perceived to exist among public officials and politicians. It is a composite index, a poll of polls, drawing on corruption-related data from expert and business surveys carried out by a variety of independent and reputable institutions. The CPI reflects views from around the world, including those of experts who are living in the countries evaluated. The Corruption Perceptions Index scores have a theoretical range between 0.0 (perceived as highly corrupt) and 10.0 (perceived as very clean). The nineteen countries we study have Corruption Perceptions Index scores ranging between 3.7 (Poland) and 9.6 (Finland), close to or better than the world's average performance (4.1) which corresponds also to the 66.6 percentile, as 1/3 of the countries score higher. Finland, Iceland and New Zealand are the world's top performing countries. The correlation between the Corruption Perceptions Index scores and the ACCI scores is high (- 0.840), particularly in the relationship with the dimension of Civil Society (political non-governmental action) and then with the dimension of Community Life.

Per capita GDP[4]

The correlation with GDP per capita (measured in PPP US Dollars) is also high (0.79) and even higher when considering the connection to the dimension of Civil Society (0.83); it is still high for Community Life participation (0.75). However, the correlation is quite low when compared to the dimension of Values (0.30). It should be noted that it is the level of per capita GDP that matters rather than its distribution, given that the correlation between the ACCI and the Gini index is below 0.4 for all the dimensions considered. This raises a number of challenging issues for future research.

There might well be some kind of Kuznets' curve for citizenship, also linked to Maslow's hierarchy of needs, implying a lower level of citizenship at early stages of development, a positive relationship between active citizenship and GDP per capita up to a certain point at which, due to the improved economic situation, citizenship stabilizes. Citizenship might decline at a later stage of development due to other factors like economic anxiety about loss of jobs or fear of globalisation.

---

[3] http://www.transparency.org/policy_research/surveys_indices/cpi

[4] Source World Bank http://www.worldbank.org/

Human Development Index[5]

The Human Development Index (HDI) can be thought of as a measure of well-being as well as a measure of the impact of economic policies on quality of life. It includes comparative measures of life expectancy, literacy, education, and standards of living for countries worldwide, ranking them on a scale ranging between 1 and 0. GDP per capita is one component of the HDI. The index was developed in 1990 by the economist Mahbub ul Haq and has been used since 1993 by the United Nations Development Programme in its annual Human Development Report (http://hdr.undp.org/reports/). The link with active citizenship can be found in the Human Development Report itself (UNDP, 2004, p. 6):

> Human development requires more than health, education, a decent standard of living and political freedom. People's cultural identities must be recognized and accommodated by the state, and people must be free to express these identities without being discriminated against in other aspects of their lives. In short: cultural liberty is a human right and an important aspect of human development—and thus worthy of state action and attention.

Table 18 shows a high and significant correlation between the HDI and the ACCI (0.79) and with two of its dimensions: Civil Society (0.84) and Community Life (0.71). Not surprisingly this resembles the relationship between the ACCI and GDP per capita. Thus, both results provide evidence to support the argument that high levels of prosperity are linked to high levels of civil and community participation. The direction of this causal link is, however, difficult to determine.

The absence of time series data prevent any statistical testing on causality. Moreover, the fundamental difficulty in establishing causal links resides in the inherent complexity of phenomena like active citizenship and the feedback and reinforcements that these variables have. On the other hand, the strong correlation found with the TICI also points to the existence of more general "enabling factors," such as respect for the rule of law, trust and attention to the common good, such as providing a developed welfare system.

Worthy of mention is the fact that both Values and Political Participation seem to have a weak relationship with all the indicators presented in Table 18.

---

5 http://hdr.undp.org/

Social cohesion

To the best of our knowledge the only index of social cohesion is the Social Cohesion Index (Green et al., 2003), which combines measures for general trust and trust in democracy, civic cooperation (attitudes to cheating on taxes and public transport), and violent crime. This index scores 15 countries (11 of which are also in the ACCI) without explaining the methodology used to assemble data coming from different sources. Another difference from the ACCI is the year of the dataset used (1996), which could partially explain the modest correlation found with the ACCI. Note that this correlation rises significantly if two countries (Sweden and Poland) are eliminated from the dataset due to the rise in correlation between the ACCI and civic cooperation and violent crime. The lack of disaggregated data prevents further analysis.

Gender Gap Index[6]

The Gender Gap Index was first launched in May 2005 by the World Economic Forum in an attempt to assess the size of the gender gap in 58 countries using economic, education, health and politically-based criteria (Hausmann et al., 2006). The Global Gender Gap Index 2006, the second in the series, covers over 115 economies, which comprehends over 90% of the world's population and was compiled by researchers from Harvard University, the London Business School and the World Economic Forum. The index measures gaps between men and women in four areas: economic participation and opportunity, educational attainment, health and survival and political empowerment. By quantifying differences between the sexes in access to resources or opportunities, rather than measuring absolute levels, the researchers sought to remove the impact of economic development. The Gender Gap Index scores have a theoretical range between 0.00 (perfect inequality) and 1.00 (perfect equality). The nineteen countries we study have Gender Gap Index scores ranging between 0.64 (Italy) and 0.81 (Sweden), close to or better than the world's average performance of 0.66. It is worth mentioning that only 1/3 of the 115 countries have scores greater than 0.68. Sweden is the top performing country in the entire set of 115 countries included in the Gender Gap Index.

The scores in Table 18 show that there is a statistically significant correlation between the Gender Gap Index scores and the ACCI scores (0.695). Nevertheless, at similar levels of Gender Gap there is high variation in the ACCI scores, whilst at similar levels of ACCI scores the variation in the Gender Gap scores is much lower. The spread in scores is greatest at lower levels of Gender Gap. For example, Luxembourg does far better than Hungary in active citizenship at a similar level of Gender Gap. Germany achieves much higher levels of

---

6 http://www.weforum.org/en/initiatives/gcp/Gender%20Gap/index.htm

Gender Gap than Luxembourg at a similar level of active citizenship. Four of the five Nordic countries in this study (Norway, Sweden, Denmark and Sweden) have top scores in both the ACCI and the Gender Gap, but Finland's performance in active citizenship is much lower than in the Gender Gap Index.

Education and training[7]

The ACCI displays weaker correlations with indicators on education and training. The highest correlation (0.6) is with the lifelong indicator (the percentage of the adult population aged 25 to 64 participating in education and training). The remaining benchmarks from the European Commission's Education and Training 2010 agenda reveal weaker relationships. This appears to indicate that education (as measured by the six benchmarks) is only weakly related to active citizenship at a country level. However, the high correlation with the HDI (which contains educational variables) suggests the need for further research.

3.9. Step 9: Back to details

The aim of the "step 9: back to details" is to go in depth and to investigate the concept of active citizenship further. Taking into account the nature of data which form the basis of the composite indicator, the active citizenship composite indicator can be developed at the individual level. This individual score enable a deeper investigation of the possible determinants of the phenomenon, encompassing country differences encoded through country level variables. In this way, a very clear picture can be drawn and the composite indicator can be easily and effectively used for policy making.

In this context, the aim of backing to detail is to deepen the understanding of Active Citizenship by identifying the determinants of Active Citizenship through the application of a multilevel model that examines both the individual level and national level characteristics. Hoskins and Mascherini (2009) presented a composite indicator to measure Active Citizenship based on 61 basic indicators drawn from the 2002 European Social Survey data. Following this framework, individual level analysis is carried out using socio-demographic and behavioral variables of gender, occupation, income, age, religion and use of media of active citizens. On a national level it provides an analysis of the contextual features of the country which enhance active citizenship such as; GDP, income equality, national averages of education and religious diversity. This research also enables a greater understanding of who is much less active and exhaustive results can be found in Mascherini et al 2009.

[7] http://ec.europa.eu/education/policies/2010/doc/after-council-meeting_en.pdf

The results of our statistical analysis, based on a multilevel regression model provide a clear identikit of the active citizen in Europe and the drivers of the phenomenon are identified both at the individual and at the country level.

At country level we can say that the level of active citizenship is higher in countries that are not just the richest but also the most equal in terms of income distribution and the most open in terms of religious heterogeneity. The meaning of the country specific variables clearly explains and characterizes the polarization on the north-south axis of the active citizenship phenomenon.

At the individual level we can say that the active citizen is working age, male or female, peaking between 48-64 the age of the baby boomers. They typically have a good level of education and are also active in lifelong learning. An active citizen typically has a clear idea of the importance of religious in their life and they actively attend the church. He or She lives in the countryside and has a good income. Concerning media, she or he watches TV a moderate amount of time and reads newspapers. In terms of employment she or he does not work in the labor market but is also not looking for a job which gives them the time to participate.

In reference to the previous literature we would say that this result is similar to what we would expect what is new is the relationship with lifelong learning and active citizenship a which reflect the importance of education and training in the active citizenship.

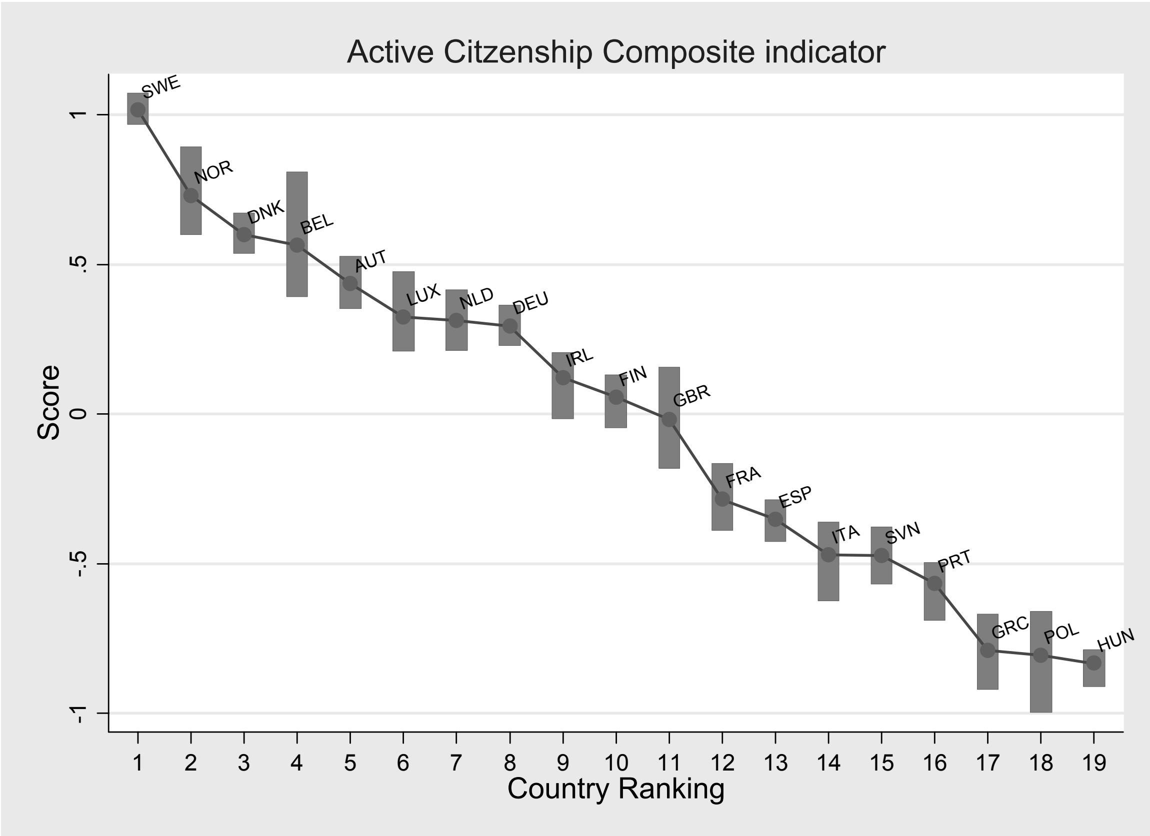
Conversely, the persons who are the least active are young people, living in big cities, with lower levels of education who are not participating in any lifelong learning activities. They typically are part of the job market but with a low income and responsibility. They have a confused view about the importance of the religion in their life and they do not attend any religious services apart from the special occasions. They spend every day a considerable amount of time in watching TV without reading any newspaper.

3.10.Step 10: Presentation and Dissemination

The scores of the countries for the ACCI are given in the tables provided above. In addition to these, the scores for each country were also reported with their associated confidence bounds, resulting from applying uncertainty analysis (see Figure 14).

Concerning the dissemination strategy, the preliminary version of ACCI was completed by CRELL in December 2006. With the availability of data from experts' survey on weights in March 2007, the ACCI was revised and finalized on October 2007.

Figure 14: Results from the uncertainty analysis of the ACCI



The composite indicator was presented at the Directorate General of Education and Culture (Brussels, December 2007) where its importance in the light of Lisbon 2010 objective was discussed and emphasized. Moreover, ACCI was presented to the scientific community in June 2008 at the 3MC conference organized by GESIS-ZUMA in Berlin, at the Mediterranean Conference on Social Science in Malta on March 2009, at the European Social Research Association Conference (Warsaw, June 2009), and finally at the International Society on Quality of Life Conference (Florence, July 2009).

The preliminary findings on ACCI have been published as EUR Report in December 2006 (Hoskins et al, 2006). Then, the final version of the composite indicator was submitted to the journal Social Indicators Research. The paper was printed in February 2009 (Hoskins and Mascherini, 2009).

Further direction of the research are pointing to the analysis of Active Citizenship and Human Values, measured with the Schwarz scale and included in the European Social Survey core questionnaire (Schwartz , 1992). This will give the final picture of the phenomenon of Active Citizenship at the European Level.

4. Summary and challenges

From Saltelli et al., 2007 it is obvious that, *<<from a purely mathematical point of view, the aggregation convention used for composite indicators deal with the classical conflictual situation tackled in multi-criteria evaluation. Thus, the use of a multi-criterion framework for composite indicators in general, and for sustainability and well-being indices in particular, is relevant and desirable (Funtowicz et al., 2002; Munda, 1997 , 2005, Ülengin et al., 2001). However, the so-called "multi-criterion problem" can be solved by means of a variety of mathematical approaches, all of them plausible. This situation is due to Arrow's impossibility theorem (Arrow, 1963), which proves that it is impossible to develop a "perfect" mathematical aggregation convention. This implies that it is desirable to have mathematical algorithms that may be recommended on some theoretical and empirical grounds, or alternatively to test how robust results are with respect to different aggregation procedures. This makes sensitivity analysis a fundamental step during the development of any composite indicator (Saisana et al., 2005; Saltelli et al., 2004).>>.*

Munda (2004) insists on the importance of the quality of the aggregation convention, which is to guarantee the consistency between the assumptions made and the ranking obtained. Such quality depends crucially on the way this mathematical model is embedded in the social, political and technical structuring process. In other words, composite indicators are context-dependent and present both technical and socio-political uncertainty.

These uncertainties may not be neglected. They must be dealt with caution and above all be made as transparent as possible. Transparency must remain one of the main ingredients in developing composite indicators. This is especially true for the problem of weighting individual indicators, which remains the most important source of uncertainty and debate.

One bad example

Among the list of objections to the use of composite indicators one reads (Saisana and Tarantola, 2002, Nardo et al., 2008):

- *Composite indicators may send misleading, non-robust policy messages if they are poorly constructed or misinterpreted [… or] may invite politicians to draw simplistic policy conclusions […]*

An example of poorly constructed composite indicator appears in the Financial Times of September, 13th 2007. The article is about an 18-country composite indicator of "sustainability of fiscal and ecological development", developed at Germany's Allianz

Insurance and Dresdner Bank, showing an unexpected result. Russia is outpacing United States, United Kingdom and Germany in securing its population's long-term economic and environmental future.

The index is a composite of five indicators, and Russia holds the top position in three of them - based on its good performance on public debt, current account and net borrowing balances, compared with the other more established economies. However, Russia performs badly on ecological development, being pushed to the bottom rank for the other two indicators because of its high carbon dioxide emissions and high energy use per unit of gross domestic product. But this is largely outweighed by the strong fiscal record.

The result came about because of Russia's huge oil and gas reserves and the sharp rise in energy prices in recent years, which have boosted significantly the country's finances. As soon as the oil price falls, then so will the country's fiscal performance. This is clearly an example of poorly constructed composite indicator as the sustainability of fiscal development of a country should not depend on the volatility of the oil price on the market. More attention should have been paid to produce a sound theoretical framework which could have guided the choice of more suitable component indicators.

In order to react to this critique, the chief economist at Allianz Insurance and Dresdner Bank declared that the findings were *<<…a thought-provoking way of looking at the issue, although in Russia's case it's probably more appropriate to look at each indicator separately>>.*

Notice here that the sole fact that a composite indicator has been put in place, whatever good or bad its architecture, will give the policy issue a great deal of visibility and will stimulate debate by concerned parties.

One good example

As stressed in Nardo et al. (2008), the strength of a composite indicator can largely depend on the availability of good quality underlying data. In July 2007, the World Bank published its sixth annual report on Worldwide Governance Indicators. The World Bank did a great effort in sifting through hundreds of indicators from 33 data sources by 30 organisations, thus providing a picture of governance across six categories over a 10-year period and more than 200 countries and territories. Given that there is no objective scale for judging a country's "rule of law", the Bank relies on up to 19 different sources for that measure.

The resulting report aimed business groups, reformists and other parts of civil society to push for better governance, which is crucial for development. The FT already expected some critics: *<<…the niceties [of CIs] are often lost when ranking countries or providing a global* *top or bottom 10 countries>>.* Nevertheless, a word of praise was given *<<…the inevitable talk of "hit parades" should not detract [the Bank] from its overall effort>>.*

To conclude this example, we cite here another sentence from the same FT article asserting that the main reason behind composite indicators is in their ability to quantify multidimensional and complex concepts even with all the limitations that their use implies: *<<Economists are often accused, justly, of thinking that what cannot be counted does not count. In this case, economists are trying to count what - many would say - cannot be counted. The alternatives, however, are worse. Either we ignore this fact or we make subjective guesses. For all its weaknesses, the Bank remains best-equipped to crunch the numbers and deliver the judgment, however unpalatable.>>*

Factors affecting country rankings

Although other factors may have a certain impact on the final scores and rankings (such as the imputation of missing values, the type of hierarchical structure chosen to represent the framework, or the aggregation method chosen), the issue of weighting plays a central role to the development of a composite indicator. The weighting, i.e. the need to combine in a meaningful way the underlying individual indicators, should ideally be made explicit (this is almost never the case for composite indicators that appear on the media) and agreed by an as-wide-as-possible public. This agreement is practically impossible to achieve, so specific participatory techniques (such as budget allocation) have been developed and can be used to take into account the different subjective value judgments without the need to agree on a unique set of weights.

In this publication we show a case study using the most straightforward of the participatory methods for weighting: the *budget allocation process*. Experts are given a *budget* of e.g. 100 points, to be distributed over a number of individual indicators. They can *pay* more for those indicators whose importance they want to stress and less or nothing for those they consider less important (Jesinghaus, 1997). The budget allocation is feasible until a maximum of 10-12 indicators. If too many indicators are used, this method can give serious cognitive stress to the experts who are asked to allocate the budget.

We strongly suggest testing the robustness of the country rankings to the weights proposed by different actors, to see whether the differences in the weights provide consistent results in terms of rankings. We will show this in our case study.

Some additional thoughts on robustness tests

Uncertainty and sensitivity analysis are key tools to investigate the robustness of the message conveyed by a composite indicator. Indeed, the building of composite indicators involves stages where judgments have to be made: the choice of a conceptual model, the selection of individual indicators, the weighting of indicators, the treatment of missing values, the aggregation rule, etc. All these sources of subjective judgment can influence the message brought by the composite indicator in a way that deserve analysis and corroboration. A combination of uncertainty and sensitivity analysis (respectively UA and SA) can help to gauge the robustness of the composite indicator, to increase its transparency and to help framing a debate around it.

Despite a synergistic use of UA and SA has proven to be more powerful (Saisana et al., 2005 and Tarantola et al., 2000), UA is more often adopted than SA (Jamison and Sandbu, 2001) and the two types of analysis are often treated separately. The types of questions for which an answer is sought via the application of UA & SA are:

• Does the use of one development strategy versus another in building the CI provide actually a partial picture of the countries' performance?

• Which constituents (e.g. countries) have large uncertainty bounds in their rank (volatile countries)?

• Which are the factors that affect the countries' rankings?

All things considered, a careful analysis of the uncertainties included in the development of a CI can render its building more robust. A plurality of methods (all with their implications) should be initially considered, because no model (CI development strategy) is a priori better than another, provided that internal coherence is always assured, as each model serves different interests. The CI is no longer a magic number corresponding to crisp data treatment, weighting set or aggregation method, but reflects uncertainty and ambiguity in a more transparent and defensible fashion.

References

Arrow, K.J. (1963) Social choice and individual values. 2nd edition, Wiley, New York

Cherchye, L., Moesen W. and Van Puyenbroeck T. (2004), Legitimately Diverse, Yet Comparable: on Synthesising Social Inclusion Performance in the EU, Journal of Common Market Studies, 42: 919-955.

Dempster A.P. and Rubin D.B. (1983), Introduction (pp.3-10), in Incomplete Data in Sample Surveys (vol. 2): Theory and Bibliography (Madow W.G., Olkin I. and Rubin D.B., eds.) New York: Academic Press.

Ebert U. and Welsch H. (2004), Meaningful environmental indices: a social choice approach, Journal of Environmental Economics and Management, Vol. 47: 270-283.

European Commission (2001a), Summary Innovation Index, DG ENTR, Brussels.

European Commission (2001b), Internal Market Scoreboard, DG MARKT, Brussels.

European Commission (2004), Economic Sentiment Indicator, DG ECFIN, Brussels,

Fagerberg J. (2001), in Lundvall B. and Archibugi D. (eds.) Europe at the crossroads: The challenge from innovation-based growth in the Globalising Learning Economy, Oxford Press.

Freudenberg M. (2003), Composite indicators of country performance: a critical assessment, OECD, Paris.

Funtowicz, S.O., Martinez-Alier, J., Munda, G., Ravetz, J. (2002) Multicriteria-based environmental policy, in H. Abaza and A. Baranzini (eds.) Implementing sustainable development. UNEP/Edward Elgar, Cheltenham, pp 53-77

Gall M. (2007) Indices of social vulnerability to natural hazards: A comparative evaluation, PhD dissertation, Department of Geography, University of South Carolina.

Green A., J. Preston and R. Sabates (2003), Education, Equity and Social Cohesion: a Distributional Model, WBL Research Report N° 7, Wider Benefits of Learning Centre, London, United Kingdom

Hammond A, Adriaanse A, Rodenburg E, Bryant D, Woodward R (1995) Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development, Washington: World Resources Institute

Hausmann R., Tyson L.D., Zahidi S. (2006) The Global Gender Gap Report 2006. World Economic Forum.

Hoskins B. and M. Mascherini (2009) Measuring Active Citizenship through the Development of a Composite Indicator, Social Indicators Research, 90(3), 459-488

Hoskins, B. (2006). Draft framework on indicators for Active Citizenship. Ispra: CRELL.

Hoskins, B., Jesinghaus, J., Mascherini, M., Munda, G., Nardo, M., Saisana, M., Van Nijlen, D., Vidoni, D., & Villalba, E. (2006). Measuring Active Citizenship in Europe, European Commission EUR 22530 EN.

Jacobs R., Smith P. and Goddard M. (2004), Measuring performance: an examination of composite performance indicators, Centre for Health Economics, Technical Paper Series 29.

Jamison, D. and Sandbu, M. (2001) WHO ranking of health system performance. Science, 293, 1595-1596.

Jencks S.F., Huff E.D. and Cuerdon T. (2003), Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001, Journal of the American Medical Association, 289(3): 305-12.

Jesinghaus, J. (1997) Current approaches to valuation. Sustainability Indicators: a Report on the Project on Indicators of Sustainable Development. B. Moldan, S. Billharz and R. Matravers. New York, John Wiley & Sons on behalf of the Scientific Committee on Problems of the Environment (SCOPE): 84-91

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). Judgments under uncertainty: Heuristics and biases. Cambridge University Press.

Kim, J., & Mueller, C. W. (1978a). Introduction to factor analysis: what it is and how to do it. Beverly Hills: Sage.

Kim, J., & Mueller, C. W. (1978b). Factor analysis: statistical methods and practical issues. Beverly Hills: Sage.

Little R.J.A. and Rubin D.B. (2002), Statistical Analysis with Missing Data, Wiley Interscience, J. Wiley & Sons, Hoboken, New Jersey.

Marshall, T. (1950). Citizenship and social class and other essays. Cambridge: Cambridge University Press.

Mascherini M., B. Hoskins and A. Manca (2009) The characterization of Active Citizenship In Europe, European Commission EUR JRC47543 EN

Melyn W. and Moesen W.W. (1991), Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available, Public Economic research Paper 17, CES, KU Leuven.

Miller M.B. (1995), Coefficient Alpha: a basic introduction from the perspectives of classical test theory and structural equation modelling, Structural Equation Modelling, 2, 3: 255-273.

Munda G. and Nardo M. (2009), Non-compensatory/Non-Linear composite indicators for ranking countries: a defensible setting, Applied Economics (41)12, 1513-1523

Munda, G. (2005) “Measuring sustainability”: a multi-criterion framework. Environment, Development and Sustainability Vol 7, No. 1, pp. 117-134

Munda G. and Nardo M. (2005), Constructing Consistent Composite Indicators: the Issue of Weights, EUR 21834 EN, Joint Research Centre, Ispra.

Munda, G. (2004) “Social multi-criteria evaluation (SMCE)”: methodological foundations and operational consequences. European Journal of Operational Research, vol. 158/3, pp. 662-677.

Munda, G. (1997) Multicriteria evaluation as a multidimensional approach to welfare measurement, in J. van den Bergh and J. van der Straaten (eds.), Economy and ecosystems in change: analytical and historical approaches. Edward Elgar, Cheltenham, pp. 96-115

Munda G. (1995), Multicriteria evaluation in a fuzzy environment, Physica-Verlag, Contributions to Economics Series, Heidelberg.

Nardo, M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman and E. Giovannini (2008) Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD publication Code: 302008251E1.

Nilsson R. (2000), Confidence Indicators and Composite Indicator”, CIRET conference, Paris, 10-14 October

Parker J. (1991), Environmental reporting and environmental indices, PhD Dissertation, Cambridge, UK.

Saisana, M., Tarantola, S., Saltelli, A. (2005) Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society A, 168(2), 307-323.

Saisana M. and S. Tarantola (2002) State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development, EUR 20408 EN

Saltelli, A. (2007) Composite indicators between analysis and advocacy, Social Indicators Research, 81, 65-77

Saltelli, A., Chan, K., & Scott, M. (2000). Sensitivity analysis. Probability and statistics series. Wiley.

Saltelli, A. Tarantola, S., Campolongo, F. and Ratto, M. (2004) Sensitivity Analysis in Practice: a Guide to Assessing Scientific Models. Wiley.

Saltelli, A., J. Jesinghaus and G. Munda (2007) Well Being Stories, Beyond GDP Conference, Experts Workshop, 19-20 November, 2007

Schwartz S.H. (1992), Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In Zanna P. (eds). Advanced in the experimental social psychology. 25. San Diego Acedemic Press, 1-65

Seaver, D. A. (1978). Assessing probabilities with multiple individuals: Group interaction versus mathematical aggregation. Tech. Report SSRI-78-3, Social Science Research Institute, University of Southern California, Los Angeles.

Stevens, J. (1986). Applied multivariate statistics for the social sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tarantola, S., Jesinghaus, J. and Puolamaa, M. (2000) Global sensitivity analysis: a quality assurance tool in environmental policy modelling. In Sensitivity Analysis (eds A. Saltelli, K. Chan, M. Scott) pp. 385-397. Wiley.

Tarantola S., Saisana M., Saltelli A., Schmiedel F. and Leapman N. (2002), Statistical techniques and participatory approaches for the composition of the European Internal Market Index 1992-2001, EUR 20547 EN, European Commission: JRC-Italy.

Tarantola S., Liska R., Saltelli A., Leapman N., Grant C. (2004), The Internal Market Index 2004, EUR 21274EN, European Commission: JRC-Italy

Trufte E.R. (2001), The Visual Display of Quantitative Information. Graphic Press, Connecticut, USA, 2nd edition (first edition 1981).

Ülengin, B., Ülengin, F., Güvenç, Ü (2001) A multidimensional approach to urban quality of life: the case of Istanbul. European Journal of Operational Research, 130, pp. 361-374

UNDP (2004). Human Development report 2004. Retrieved from: http://www.undp.org.in/hdr2004/HDR2004_complt.pdf

Verba, S., & Nie, H. (1972). Participation in America: Political democracy and social equality. New York: Harper and Row.

Erakunde autonomiaduna
Organismo Autónomo del

EUSKO JAURLARITZA
GOBIERNO VASCO

Eustat
EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADÍSTICA