Well Being Stories

Andrea Saltelli, Jochen Jesinghaus and Giuseppe Munda

European Commission, Joint Research Centre, Ispra (Italy) [1]

Contribution to the Experts Workshop at the
Beyond GDP Conference, EU Parliament, Brussels 19-20 November 2007

andrea.saltelli@jrc.it

http://composite-indicators.jrc.ec.europa.eu & http://esl.jrc.it/dc

Abstract

This paper deals with the difficult issue of measuring well-being. We are aware that this concept is very complex and we are not giving any final answer here. However we are convinced that even the simple attempt of measuring well-being may produce valuable lessons to be learned at a societal level. This paper addresses three main issues:

1. *The role of GDP as main policy indicator;*
2. *Lessons learned from already existing well-being stories;*
3. *Practical steps to develop an empirical well-being composite indicator.*

1 Disclaimer: The opinions are those of the authors and not of the European Commission

GDP and beyond. An introduction

The debate on the misuse of GDP as an indicator of well-being is almost as old as GDP itself; "*The welfare of a nation can scarcely be inferred from . . . [the Gross Domestic Product]."* The quote is from Simon Kuznets, the inventor of America's national-income accounts, who was eminently sceptical of the possibility of using GDP as a measure of overall well-being."[1] What Nobel Prize winner Simon Kuznets said in 1934, over 70 years ago, has still not led to a better solution for the observable misuse, in media and politics, of GDP as a proxy for well-being. In recent years, a growing stock of literature has been written about well-being, as well as on analytic problems associate to GDP, which we do not review here[2]. The debate has also invested the OECD and the European Commission, which devoted a number of recent conferences to the issue of well-being or happiness in the framework of the "Measuring Progress" framework[3].

A relevant question is hence the following: what are the perspectives to build on the existing knowledge and consensus alternative measures of well being?

Both broad and narrow definitions of well-being exist. A problem of broad definitions of well-being – in relation to available statistical information – is that broad definitions would provide little guidance on the selection of relevant indicators, since almost all indicators collected by statistical offices and international agencies are – must be – of some relevance to the well being of citizens. At the same time narrow definitions of well being tend to lack key characteristics of concern to policy. An open problem for a well being research agenda is then how to set up a framework which is both theoretically defensible and operationally useful.

A starting point is to connect well-being research with the sustainability agenda. This allows us to draw upon results already established in the literature and widely accepted by the political and scientific communities. One shared non-controversial result of the sustainability literature is that sustainability is a multidimensional concept, which should at least include economic, social, environmental and institutional dimensions.

Taking sustainability into account creates in fact a need for the inclusion of the physical appraisal of the environmental impact on the socio-economic system. Systemic approaches to sustainability issues consider the relationships between four systems: the economic, the human, the political and the natural systems.

- The *economic system* includes the economic activities of humans, such as production, exchange and consumption. Given the scarcity posit of meta-economics, such a system is efficiency-oriented.

- The *human system* comprises all activities of human beings on our planet. It includes the spheres of biological human elements, of inspiration, of aesthetics, and of morality, which constitute the frame of human life.

- With growing globalisation, collective aspects of human activities become more important; Agenda 21 captures this growing societal complexity under the heading "institutions", while in recent years many refer to "governance". We chose to call this sphere the *political system*. Since it is clear that the economic system does not constitute the entire human system, one may assume that the economic system is a subsystem of the human system.

- Finally, the *natural system* includes both the human system and the economic system. This is clearly true for well-being, too:
  Can anyone be happy living in a contaminated area?
  Without money?
  Without democratic representation?
  Witnessing deprivation of other human beings?

The next point to look at is whether there is a multidimensional measurement framework able to cope with all these issues simultaneously. *From an economic point of view*, Gross Domestic Product (GDP) has been traditionally considered as the best performance indicator for measuring national economy and welfare. But if resource depletion and degradation are factored into economic trends, what emerges is a radically different picture from the one depicted by conventional methods. A point of scientific controversy present in the debate on sustainability measure is the use of monetary or physical indexes. Examples of monetary indexes are Daly and Cobb (1989) ISEW (Index of Sustainable Economic Welfare)[4], Pearce and Atkinson (1993) Weak Sustainability Index[5], the so-called El Serafy approach[6]. Examples of physical indexes are HANPP[7] (Human Appropriation of Net Primary Production, the Ecological Footprint[8], MIPS[9] (Material Input Per unit of Service).

Although these approaches may look different, they all have one common characteristic:[10]

- These indexes are based on the hypothesis that a common measurement rod needs to be established for aggregation purposes (e.g. variables expressing money, energy, space, and so on). This creates the need to make very strong assumptions on conversion coefficients to be used and on the acceptable degree of compensability (e.g. until which point better economic performance may be justified at the expense of environmental destruction or social exclusion?). The mathematical aggregation convention behind an index thus needs an explicit and well-thought formulation.

- These indexes are somewhat confusing if one wishes to derive policy suggestions. For example, by looking at ISEW, we could know that a country has a worse sustainability performance than the one pictured by the standard GDP, but so what? Since ISEW is so aggregated, it does not provide us with any clear information on the cause of this bad performance, and it is thus useless for policy-making (while conventional GDP is at least giving clear information on the economic performance).

- The same applies to the Ecological Footprint, which sometimes can even give misleading policy suggestions; given that diet is used, it would imply that a more energy intensive agriculture might reduce the Ecological Footprint of e.g. a city, but in reality – if waste is factored in - its environmental performance would be much worse! Likewise for the weak sustainability index, which is nothing else but the classical golden rule of growth theory, where environmental physical destruction is never considered – above all if it is externalised outside the national borders.

- Last but not least, while using a single metric gives the impression of procedural scientificity, it obfuscate a possible negotiation about weights, which are in theory absent but in practice implicit in the assumption used for the conversion. Explicit weights would go in the direction of transparency.

All these approaches belong to the more general family of composite indicators and as a consequence, some assumptions used for their construction are common to them all. Notwithstanding the limits just mentioned, a conclusion that we can borrow from the sustainability literature is that composite indicators could be an adequate approach to measure overall performance regarding multidimensional concepts such as sustainability or well-being provided the temptation of a single metric is resisted and that a process of negotiation can be set in motion around these measures.

Composite indicators are very common in fields such as economic and business statistics and a variety of policy domains such as industrial competitiveness, sustainable development, globalisation and innovation[11]. The proliferation of this kind of indicators is a clear symptom of their political importance and operational relevance in decision-making[12].

From a purely mathematical point of view, it is obvious that 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 indexes in particular is relevant and desirable (Funtowicz et al., 2002[13]; Munda, 1997[14], 2005[15], Ülengin et al., 2001[16]). 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[17], 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[18]; Saltelli et al., 2004[19]).

The quality of the aggregation convention is, yet, an important ingredient to guarantee the consistency between the assumptions made and the ranking obtained. Indeed, the overall quality of a composite indicator depends crucially on the way this mathematical model is embedded in the social, political and technical structuring process (Munda, 2004[20]). This is especially true for the choice of weights that remains the most important source of uncertainty and debate. Weights are and (according to our opinion) must be *context-dependent* since they reflect political, social and economic priorities and depend on the development path a country (or a group of countries, as in the case of the European Union) wants to pursue. Policymakers may use "science" instrumentally to disguise lobbies, individual interests, or incompetence. Again, whatever aggregation procedure is used, transparency must remain one of the main ingredients: without *transparency* the *interpretability* and *coherence* of composite indicators are difficult to achieve.

From the above discussion the following main conclusions can be drawn:

- Well-being and sustainability are both multidimensional concepts whose main characteristics are very similar.

- Composite indicators may provide an adequate measurement framework for a multidimensional concept such as well-being.

These conclusions may constitute the key principles for the construction of a possible well-being composite indicator. However, one should remember that 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. In the next Section some well-being stories will be examined to highlight possible fallacies in the construction/use of aggregated measures. A later section will suggest some practical steps that may lead towards the possible construction of a well-being composite indicator.

Some Well-being Stories

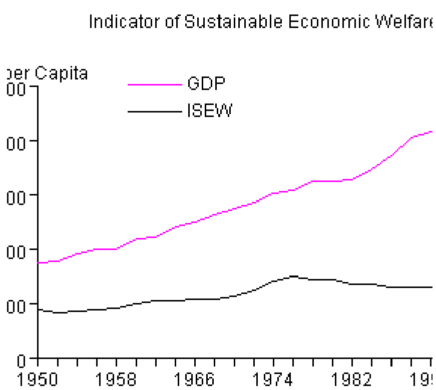
First story – The index of sustainability of fiscal and ecological development

"Russia is outpacing the US, UK and Germany in securing its population's long-term economic and environmental future, according to a new study".

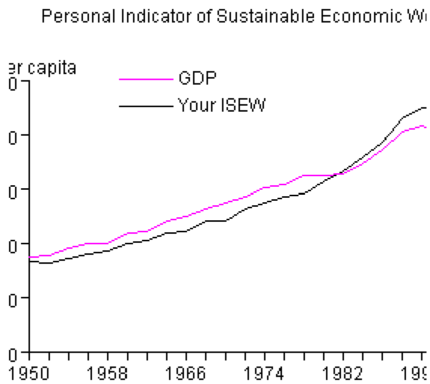
The headline of a Financial Times (September 13 2007) article looking with polite disbelief at an 18-country index of "Sustainability of fiscal and ecological development" developed by Economists at Germany's Allianz Insurance and Dresdner Bank. The index ranks Russia sixth in, ahead of the UK (placed seventh), Germany (ninth), and the US (17th).

The index is a composite of five indicators: public debt, current account and net borrowing balances, carbon dioxide emissions and energy use per unit of gross domestic product. Initial skepticism may incite us to have a closer look at the methodology, and the discovery that the massive transfer of wealth from oil-importing to oil-exporting countries caused by steadily rising energy prices shows first unexpected geopolitical shifts – expressed here as "current account balance".[21]

Second story – The Index of Sustainable Economic Welfare, ISEW



The two following graphs have been drawn using the "Create your own ISEW" feature of the Friends of the Earth website. Here is the UK ISEW as we all know it: GDP is rising steadily, while the ISEW reaches ist peak in 1975, shortly after the First Oil Crisis (1973/74), and shortly after the publication of "Limits to Growth" (Meadows & Meadows & Randers, 1972) – the large gap shown here is The Scientific Proof that we have already exceeded the environmental limits, and that we urgently need to change course to Save The Planet.



Here is a slightly modified purely environmental ISEW: we eliminated all "social", i.e. inequality, Gini, household work etc. corrections but kept the environmental, i.e. air pollution and climate change items. The result: GDP is still rising steadily, but the ISEW outperforms GDP from 1982 onwards! No surprise for those who like studying methodologies, and are aware of the partial success of environmental policies in the last decades, but it is not exactly the message that the ISEW (and Genuine Progress Indicator, GPI) authors want to send…

Third story - The Ecological Footprint

Carrying capacity of ecosystems including humans are difficult to compute as humans' footprint depends on population density, consumption levels as well as on technology[22]. What are the policy implications of the remark that Netherlands occupies 15 times its actual size in E-footprint? Would the impact of Netherlands' people for the world be better if they were spread thinly over free spots on the planet? How would the EF capture the practice of exporting production and pollution abroad? Can we trust EF data when we know that the absorptive capacity of trees depends on age and location and that world averages are used? These are but some of the most common criticism to EF heard in Ecological Economics circles.

The Ecological Footprint, while an excellent advocacy tool, suffers from the same weakness as ISEW, GPI etc., i.e. from the attempt to aggregate disjoint indicators on the basis of "acres". Neither dollars nor acres are neutral enough to be applied to a wealth of indicators that do not always fit neatly into the metaphor (try using "acres" for the fragmentation of landscapes, or to subtract "dollars" from GDP for correcting gender inequality). The EF is indeed a strong narrative underpinned by a weak model.

Some inference from the Well-being stories

First suggested inference: The metric fallacy

In our opinion ISEW and the Ecological Footprint suffer from the same problem:

"*Incommensurability, i.e. the absence of a common unit of measurement across plural values, entails the rejection not just of monetary reductionism but also any physical reductionism (e.g. eco-energetic valuation). However it does not imply incomparability. It allows that different options are weakly comparable, that is comparable without recourse to a single type of value*" [23]

The same considerations apply to other attempts to 'adjust' GDP, e.g. by satellite accounts. The latter are an excellent tool for modelling the interactions between the economy and the environment. Satellite Accounts are an essential tool for energy use, greenhouse gases, and a handful of other indicators; however, we fear that non-practitioners misunderstand “environmental accounts” as a generally applicable methodology for calculating a “Green GDP”.

Second suggested inference: The danger of reductionisms - only economy and environment matter

As the stories of the index of the Allianz-Dresdner Bank (sustainability of fiscal and ecological development) and the Ecological Footprint show, reducing well-being (in the sense of Eudaimonia) to economy and environment runs two major risks:

- to yield a measure scarcely informative for both dimensions. Which implications for policy can Russia draw from the Allianz-Dresdner Bank index?

- to leave an entire universe of dimensions uncharted. Would anyone like to live in a country which is prosperous and ecologically minded under a dictator? Would we accept therein the existence of slaves? (Aristotile would!). How can we build a measure of well-being in dollars or acres forgetful of equity, cohesion, education and culture? Who would be convinced of it across disciplines and in society?

The environment has an important role in politics, but assuming that economists (including the National Accountants whose GDP some want to modify), social scientists and others can be treated as “optionals”, is a tactical error that will not promote the good cause of the green community that was so active at the “Beyond GDP” conference.

Third suggested inference: Where to go

There is a long path ahead toward a model for well-being - while we do not know what the final model will be, we know that the initial input variable set will include as a point of departure what statisticians have painfully collected.

A look at EUROSTAT’s “key” indicators shows that the environment does indeed play a significant role among the 250 Sustainable Development and Lisbon (Structural) indicators present there. But any attempt to go “Beyond GDP” must recognise the complexity of the task to measure societal progress (well-being, sustainability, Eudaimonia or whatever label one wants to adopt). Picking a dozen of environmental variables, adding a handful of economic and social fig leaves, and aggregating them on the basis of some magic but unfortunately not so transparent unit is clearly not bound to be a successful strategy when trying to challenge GDP growth in its role as key policy guidance indicator.

Headline (composite) indicators?

By the way, NOT aggregating a battery of indicators is also not a viable option; of course, nobody seriously expects citizens, journalists or decision-makers to discuss politics on the basis of 250+ “key” indicators. But even if we limit the set to the “headline” indicators proposed, it is difficult to see how GDP could be challenged. Neither the 13 Sustainable Development headline indicators, nor the 14 Lisbon “headlines” are sufficiently simple to be easily understood by the general public – and too many for making *one* headline. So when in 2004 the Commission published the 14 ‘Structural’ headline indicators, the Financial Times produced itself the aggregation to slam the story in page 3[24]! In addition the choice of headline indicators, although the result of lengthy consultations between EU expert groups, may lead to unexpected results. If the Financial Times had aggregated, instead of the Lisbon indicators, the 13 Sustainable Development headline indicators, FT readers might have been perhaps surprised to discover that Romania is the most sustainable country in the European Union.

Headline indicators, **if successful in the media**, would have one dangerous consequence: Splitting the key indicators into a visible "A" and an invisible "B" league implies that politicians would be rewarded, by journalists discussing their merits, only for actions and spending aimed to improve "A" league headline indicators. A rationally acting politician would thus convene two meetings for discussing the yearly budget allocation: One for a handful of "A" league stake-holders (where concrete decisions on budgets would be taken), and a second one for a big crowd of "B" league have-nots, where the Minister would solemnly declare her/his good intentions to advance the valuable issues discussed here, in spite of budget constraints that are, unfortunately, not under the control of the Minister.

Such clearly unacceptable but inevitable side effects could be avoided if the headlines were composite indicators: In this case, the "A" league stake-holders would represent the interests of all those whose interests are measured by the detailed and complete underlying indicator set – therefore no need for a second "solemn declarations only" meeting, and no policy distortion.

A carefully modelled aggregation performed by respected academic or international organisations can bring to page 1 of the literate press issues as diverse and as specialised such as university rankings, good governance, biodiversity, and the freedom of press, which would otherwise be lost on page 14.

Here what the Financial Times says about the World Bank sixth annual Worldwide Governance Indicators (July 12):

"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."

A research path

There are already a number of initiatives attempting to quantify "wealth" in the broadest sense: *Genuine Savings* of the World Bank, the *Redefining Progress Genuine Progress Indicator* (GPI) and the closely related *Index of Sustainable Economic Welfare*, the *Canadian Index of Well-being* (CIW), Robert Prescott-Allen's *Human Well-being Index*, and the *New Economics Foundation's Happy Planet Index*. As discussed earlier, although these measures have been effective in promoting the debate, none of them would be strong enough to challenge GDP growth in its role as policy guidance indicator, neither is it clear whether one such measure will ever exist and be accepted. Nevertheless the investigation is worth undertaking, and in the following, we propose some concrete steps.

Identification of indicators

According to the already mentioned Sustainable Development Strategy (SDS), as adopted by the European Council in June 2006, sustainable development 'aims at the continuous improvement of the quality of life and well-being on Earth for present and future generations'. Measuring success or failure on the road to well-being is a complex task, although we do not start from scratch. First the good news: There are indeed some objective measures showing a strong correlation to subjective well-being. For example, people living in countries with a very unequal income distribution tend to be slightly unhappier. Unemployed people are less satisfied with their lives than people with a job. Now the bad news: We have just scratched the tip of the iceberg. The Gini coefficient of income distribution and the unemployment rate are good candidates, but there is a long queue of other objective measures that might be influential for individual well-being. How can we then select some individual indicators that are scientifically sound, empirically computable and policy relevant? We propose to start with screening the wealth of existing databases for well-being indicators.

For example, EUROSTAT's Sustainable Development indicators database contains over 180 indicators, including breakdowns by gender etc.; about 55 can be categorised as economic, about 65 as social, and about 50 as environmental indicators. Some 20 indicators resist such categorisation, and might be called "political" or "institutional" indicators (using Agenda 21 language). These indicators can be extracted and the data condensed into a single Excel sheet that we propose to call the *Cube* since it presents three axes: countries (34), time (about 20 years), indicators (*many*). On a neighbouring URL, EUROSTAT hosts over 130 Structural Indicators for monitoring the Lisbon strategy – for the sake of completeness, we can add them to the Cube. There is some overlap, but overall the Cube will host well over 200 "key indicators" that are our starting point for a thorough screening for well-being candidate indicators. If we restrict the screening to the 27 European Union countries, the last ten years, and 200 indicators, then the Cube has 27*10*200=54,000 cells.

Of course, we need to explore what would be a good compromise between a scientifically correct and a feasible well-being indicator. One important assumption is the following: behind each of EUROSTAT's 200 key indicators there is an expensive data collection supported by relevant scientific and political stakeholders lobbying Member States' statistical services; declaring their favourite indicators as "not important enough" would be an act of social exclusion. So there seems to be no easy way to reduce the list of 200 indicators to a "handy" set.

Another possible approach might be using statistical techniques, thus achieving an in depth analysis of the meaning and consequence of each single indicator. Multivariate analysis is particularly useful in this context. In general, the underlying nature of the data needs to be carefully analysed 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 construction phase of the composite indicator. Information can be grouped and analysed along at least two dimensions: individual indicators and countries.

Grouping information on individual indicators. The analyst must first decide whether the nested structure of the composite indicator is well-defined and if the set of available individual indicators is sufficient or appropriate to describe the phenomenon. This decision can be based on expert opinion and the statistical structure of the data set. Different analytical approaches, such as principal components analysis, Factor analysis (or correspondence analysis which makes no distributional assumptions, see Manly 1994[25]) and Cronbach Alpha[26] 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 needed.

Grouping information on countries. Cluster analysis[27] 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 from medicine to psychiatry and archaeology. Cluster analysis is also used in developing composite indicators to group information on countries based on their similarity on different individual 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 to impute missing data with a view to decreasing the variance of the imputed values.

Designing an indicator tree structure

It is essential to acknowledge that the number of indicators to be included in the composite will be high – there is no credible way to describe and measure policy-making in the 21st century with a handful of indicators. Given that human beings have a limited *channel capacity*, i.e. a limited ability to process simultaneously distinct pieces of information, a high number of indicators is a serious obstacle to communication. Half a century ago, the psychologist George Miller[28] revealed that the human mind can handle about seven stimuli with acceptable precision. For example, an average person can distinguish about seven musical tones by their pitch. Trained musicians, however, may be able to separate up to 50-60 different pitches. The phenomenon is closely linked but not identical to short-term memory: To compare the elements of a stream of information, we need to keep each element "stored in memory" for the time required to complete the analysis. So what are the implications for the composite indicator of well-being?

To illustrate the approach, below a tree structure in which each element has no more than seven sub-elements. For example, a "tree" with 4 main themes, seven sub-themes per main theme and seven indicators per sub-theme has 4*7*7=196 indicators. For levels 1 and 2, this might look as follows:

Economy		Social issues		Environment		Politics & Culture	
1	Labour market	8	Education	15	Climate change	1	Media diversity
2	Innovation	**9**	**Health ***	16	Toxic chemicals	2	E-government
3	Investment	10	Elderly	17	Biodiversity	3	Legal system
4	Productivity	11	Children	18	Resource use	4	Active citizenship
5	Competi-tiveness	12	Minorities	19	Waste	5	Bureaucracy
6	Trade	13	Drugs & alcohol	20	Air pollution	6	Corruption
7	Government debt	14	Gender equality	21	Water pollution	7	Political stability

Each of the 28 sub-themes above would then contain 7 indicators. For example, the social indicator No. 2, "**Health***", might be composed of:

1. Overweight people
2. Health care expenditure
3. Suicide death rate
4. Present smokers
5. Serious accidents at work
6. Resistance to antibiotics
7. Incidence of salmonellosis

We have chosen the health example because most of us may have an idea of what indicators belong to “health”. For many of the 28 sub-themes, we may not feel comfortable if somebody asks us which are the right indicators to choose, or which weights we would assign to a given list of indicators of (for example) biodiversity. Experts for *biodiversity* might find it easier, however, to negotiate among themselves the most important indicators from the EUROSTAT list, and to give them tentative weights.

Weighting of individual indicators within the composite

Central to the construction of a composite indicator is the need to combine in a meaningful way different dimensions measured on different scales. This implies a decision on which weighting model to be used and which procedure to be applied to aggregate the information. Weights should ideally be selected according to an underlying and agreed, or at least clearly stated, theoretical framework. Weighting implies a “subjective” evaluation, which is particularly delicate in case of complex, interrelated and multidimensional phenomena. The menu of weighting methods is rather rich and increasing with the creativity of the practitioners. Ideally, weights should reflect the contribution of each indicator to the phenomenon measured by the overall composite, in our case: well-being.

Most composite indicators rely on equal weighting, i.e., all variables are given the same weight. This could correspond to the case in which all variables are “worth” the same in the composite. Alternatively, it could be the result of insufficient knowledge of causal relationships, or ignorance about the correct model to apply (like in the case of Environmental Sustainability Index - World economic forum, 2002), or even stem from the lack of consensus on alternative solutions (as happened with the Summary Innovation Index - European Commission, 2001). In any case, equal weighting does not mean "no weights", but explicitly: All issues measured are equally important. In some cases, it may happen that - by combining highly correlated variables - one introduces an element of double counting into the index. Moreover, if variables are grouped into dimensions and those are further aggregated into the composite, then applying equal weighting to the variables may imply an unequal weighting of the dimension (the dimensions grouping the larger number of variables will have higher weight). This could result in an unbalanced structure of the composite index.[29]

Notice that there will almost always be some positive correlation between different measures of the same aggregate. Thus, a rule of thumb might be introduced to define a threshold beyond which the correlation is a symptom of double counting. However, relating correlation analysis to weighting can be very dangerous, in particular but not only when motivated by apparent redundancy. For example, in the CI of e-business readiness the indicator I1 "Percentage of firms using Internet" and indicator I2 "The percentage of enterprises that have a web site" display a correlation of 0.88 in 2003: are we allowed to give less weight to the pair (I1, I2) given the high correlation or shall we consider the two indicators as measuring different aspects of innovation and communication technologies adoption and give them equal weight in constructing the composite indicator? If weights should ideally reflect the contribution of each indicator to the composite, double counting should not only be determined by statistical analysis but also by the analysis of the indicator itself vis à vis the rest of indicators and the phenomenon they all aim to picture.

Weights may also reflect the statistical quality of the data, thus higher weight could be assigned to statistically reliable data (data with low percentages of missing values, large coverage, sound values). In this case the concern is to reward only easy to measure and readily available indicators, punishing the information that is more problematic to identify and measure.

Statistical models such as *principal components analysis* or *factor analysis* could be used to group individual indicators (Nicoletti et al., 2000[30]). These methods account for the highest variation in the data set, using the smallest possible number of factors that reflect the underlying "statistical" dimension of the data set. Weighting only intervenes to correct for the overlapping information of two or more correlated indicators, and it is not a measure of theoretical importance of the associated indicator. Weights, however, cannot be estimated if no correlation exists between indicators. Other statistical methods, such as the *benefit of the doubt* approach (BOD, see Melyn and Mosen, 1991[31], and Cherchye et al, 2004 [32]) is extremely parsimonious about weighting assumptions, as it lets the data decide on the weights and is sensitive to national priorities.[33] BOD employs linear programming tools to estimate an efficiency frontier that would be used as a benchmark to measure the relative performance of countries. As the essence of BOD is that it yields most favourable, country-specific weights, its authors believe it may help to counteract problems that stem from the use of fixed sets of weights for all countries. For example, several European policy issues on well-being may entail an intricate balancing act between supra-national concerns and the country-specific policy priorities of member states. If one opts to compare the multi-dimensional performance of EU member states by subjecting them to a fixed set of weights, this may prevent acceptance of the entire exercise, given that each member state has its own national specificity.

Multiple regression models can handle a large number of indicators (see a standard textbook in econometrics like Green, 2000[34]). This approach can be applied in cases where the sub-indicators considered as input to the model are related to various policy actions and the output of the model is the target. The regression model, thereafter, could quantify the relative effect of each policy action on the output, i.e. the single indicator. However, this implies the existence of a "dependent variable" (not in the form of a composite indicator) that accurately and satisfactorily measures the target in question. In the case of the well-being CI, subjective happiness might be used as a dependent variable – and this is in fact one of the frequently discussed questions in happiness research. Measuring the influence of a number of independent variables on this policy target is a reasonable question. Alternatively such an approach could be used for forecasting purposes. In a more general case of multiple output indicators, *canonical correlation analysis* that is a generalization of multiple regression could be applied. However, in any case, there is always the uncertainty that the relations, captured by the regression model for a given range of inputs and outputs, may not be valid for different ranges.

Unobserved components is similar in spirit to the multiple regression models. It does not need an explicit value for the "dependent variable" as it treats it like another unknown variable to be estimated. This advantage is counter-balanced by the estimation complexity and the computational cost.

Instead of sophisticated statistical approaches, *participatory methods* that involve various stakeholders – experts, citizens and politicians – can be used to assign weights. This approach has the advantage that it is intuitively understandable to the non-statistician, and is thus more transparent to users.

The most straightforward of the participatory methods is the *budget allocation* approach (BAP): 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 in Moldan and Billharz, 1997 [35]). 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. In the tree structure designed above, each element has no more than seven sub-elements, which effectively eliminates such cognitive stress.

A special case of "budget allocation" are public opinion polls: Here, the "experts" are ordinary citizens. Polls have been extensively used over the years, as they are easy and inexpensive to carry out (Parker, 1991). However, they are limited to themes that ordinary citizens can judge, and are therefore not possible for the level of detail offered by EUROSTAT's databases. Opinion polls could and should be used for the top level of aggregation, e.g. for determining the relative weights of the economic, social, environmental and institutional pillars; at this level, citizens, i.e. voters, are by definition "experts".

The *analytic hierarchy process* (AHP, pair wise comparison of attributes, Saaty, 1987 [36]) and conjoint analysis (comparison of attributes on different levels) are also widely used techniques for multi-attribute decision making, since they enable the derivation of overall attribute (i.e. individual indicator) importance based on a number of rotating attribute comparisons, as opposed to directly assigning weights. AHP weights are less sensitive to errors of judgement, and allow for estimating the degree of inconsistency in the experts' opinion. However, since the AHP is based on pair-wise comparisons of indicators, it is applicable only to a very small number of indicators. Moreover, since AHP questions yield *ordinal* answers, their translation into *cardinal* weights has no sound theoretical foundation.

Conjoint analysis derives the worth of the individual indicators from the worth of a composite, i.e. it reverses the process of AHP, with which it shares advantages and disadvantages. Further complication is the need to specify and estimate a utility function (see Hair et al. 1995[37], and Green and Srinivasan 1978[38] among others).

Monetisation for weighting the components. A well-known technique to determine weights of a set of indicators, is monetisation. GDP, a composite indicator with around 500 component indicators, uses market prices and costs (for government activities) to calculate its "overall score"; and it has been frequently suggested to use monetary units also for issues that do not have an observable market price. Three main monetisation methodologies have been widely tested in the past, inter alia in the mid-1990s multi-country EU-funded "Externalities of Energy" (ExternE) and GreenStamp projects [39]:

- Willingness to Pay (WTP): "How many Euros per year are you willing to pay for lower CO_2 emissions?"

- Willingness to Accept (WTA): "What would you be willing to accept as a compensation for unsafe streets in your neighbourhood?"

- Observed Avoidance Cost (OAC): "How much does our society spend to lower the rate of street accident victims?" (assuming that expenditures for road safety correspond to the value we attach to human lives).

In practice, WTA yields systematically values three times higher than WTP [40], and both methods are not easily applicable to the majority of the indicators in the Sustainable Development set. The OAC (or "Sustainable National Income", SNI) approach developed originally by the CBS statistician Roefie Hueting suffers from the assumption that societal spending is based on rational thinking and risk assessment (plus a number of more technical obstacles[41]). Overall, monetisation is fraught with many methodological problems. The method has frequently been tested on the most prominent environmental indicator, CO_2 emissions, and studies published by honest scientists differ by four orders of magnitude, i.e. a ratio of 10,000:1 between the highest and the lowest CO_2 damage value. This corresponds to saying "we could not agree whether CO_2 should have, in our environment index, a weight of 99% or of 0.01%". Such huge uncertainty does not prevent, for example, the World Bank from using a CO_2 damage estimate of *exactly* 20 US$/ton of carbon in calculating "The Wealth of Nations" alias Genuine Savings [42]; or Redefining Progress from using 89.57 US$/ton in their Genuine Progress Indicator (GPI) [43]; but it might be unwise to use such controversial values in an official Composite Indicator of Well-being.

The basic premise of our approach is to use different weighting methods and test their impact on the final country scores (and ranking). Weights and weighting models usually have an important impact on the results especially when higher weights are assigned to individual indicators on which few countries excel or fail. This is why weights need to be made explicit and transparent. However, the impact of a weighting method should not over-emphasized. In our experience, other factors have the same or a higher impact on final scores and rankings, such as the imputation of missing values, the type of hierarchical structure chosen to represent the framework, or also the aggregation method chosen. Moreover, the reader should bear in mind that, no matter which method is used, weights are essentially value judgments and have the property to make explicit the objectives underlying the construction of a composite indicator (Jacobs et al., 2004[44]). Finally, one not so obvious "weighting" problem has by far the greatest impact on country scores: The non-inclusion of important issues, giving them effectively a weight of zero (we discussed this earlier under "headline indicators").

Again, whatever method is used to derive weights, no consensus is likely to exist. This should not preclude use of a composite, but highlights the dangers of presenting any composite as "objective". At best, it indicates a set of priorities that has been informed by popular or expert judgments (including the analyst). Assumptions and implication of the selected weighting system should always be made clear and tested for robustness. Soundness and transparency should guide the entire exercise.

Linear, non-linear and geometric weighting

The great majority of the ca. 100 known composites[45], including the popular Human Development Index (HDI), use a fairly straightforward linear weighting method:

1. All indicators are normalised with the MinMax formula, where Score=MaxPoints*(current value-worst)/(best-worst).[46]

2. Indicator scores are multiplied with their weights, and the overall score is divided by the sum of the weights (this step may be repeated for sub-indices if more than one level of aggregation is being used, see e.g. the Yale/Columbia Environmental Sustainability Index, ESI)

However, different aggregation rules are possible. Individual indicators could be summed up, multiplied or aggregated using nonlinear or geometric techniques. Each technique implies different assumptions and has specific consequences.

The commonly applied *linear normalisation and aggregation* is useful when all individual indicators have the same measurement unit and further ambiguities due to scale effects have been neutralized (for details see Nardo et all, 2005[47]). Non-linear (e.g. logarithmic) *normalisation* can be used for step 1 above for strongly skewed indicators.

In contrast, *geometric aggregations* (in which indicators are multiplied and weights appear as exponents) have been suggested in cases where non-comparable and strictly positive individual indicators are expressed in different ratio-scales. The absence of synergy or conflict effects among the indicators is quoted as a necessary condition to allow either linear or geometric aggregations. Furthermore, linear aggregations reward indicators proportionally to the weights and irrespective of the indicators' scores, while geometric aggregations reward more those countries with higher indicator scores.[48]

In both linear and geometric aggregations weights express trade-offs between indicators: the idea is that deficits in one dimension can be offset by surplus in another (Munda, and Nardo, 2007[49]). With linear aggregations, the compensability is constant, while with geometric aggregations compensability is lower when the composite contains indicators with low values. In policy terms if compensability is admitted (as in the case of pure economic indicators) a country with low values on one indicator will need much higher scores on the other indicators to improve its situation if the aggregation rule is a geometric one. Thus, in a benchmarking exercise, countries with low values would be favored by a linear rather than a geometric aggregation. On the other hand, in a geometric aggregation the marginal utility of an increase in the score would be much higher when the absolute value of the score is low. Consequently, if the aggregation is geometric, a country should be more interested in increasing those sectors/activities where its performance is relatively low in order to have the highest chance to improve its position in the ranking. In contrast, a country has interest in specializing along its most effective dimensions when the aggregation is linear (in practice, applying the logic of diminishing returns, the "most effective dimensions" will often be the weak ones anyway – little investment is needed to leave the last ranks).

When different goals are equally legitimate and important, then a non-compensatory logic may be necessary. This is usually the case when very different dimensions are involved in the composite, like in the case of sustainability indexes, where physical, social and economic figures must be aggregated. If the analyst decides that an increase in economic performance can not compensate a loss in social cohesion or a worsening in environmental performance, then neither the linear nor the geometric aggregation are suitable. Instead, a non-compensatory *multicriteria approach* will assure non compensability by formalizing the idea of finding a compromise between two or more legitimate goals[50].

5. Robustness and sensitivity

One key issue to be addressed is the use of uncertainty and sensitivity analysis in the investigation of the robustness of the message conveyed by the composite indicator. The construction of composite indicators involves in fact 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 will influence the message brought by the composite indicators 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 that a synergistic use of UA and SA has proven to be more powerful (Saisana et al., 2005[51] and Tarantola et al., 2000[52]), UA is more often adopted than SA (Jamison and Sandbu, 2001[53]) 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 construction 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 construction 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.

Conclusions

From a purely technical point of view, it is easy to calculate a comprehensive overall Well-Being Indicator, givenfor example, the 250+ indicators in EUROSTAT's databases, and the knowledge and the tools to perform the aggregation. What is difficult is to do that in a transparent, scientifically credible, participatory, consensual and politically defensible way.

In this brief note we have described one possible route to discuss well-being: the cautious step-wise construction of composite indicators built on a sufficient number of individual statistical variables.

We are led to this route by several considerations:

- our reservation toward the metric fallacy, e.g. the temptation to reduce the well-being issue to a single metric based on some fancy common unit, be it metres or acres or tonnes or euros;

- the observation that context dependent weights allow our societies to negotiate on what counts for the well-being of the citizens in a transparent manner;

- our belief in the capacity of composite indicators to capture the interest of democratic constituencies via the impact that these measure have on the media;

We are aware of the wealth of available statistical information available in the EU and puzzled by its parsimonious use. Each of the 250 or so indicators held at EUROSTAT is "key" to society, and took many years to be defined, agreed upon, and collected. We are convinced that a good deal of what is relevant to the well-being of the citizens is already contained in these data.

Hence even if we do not know yet what model could finally emerge if these variables were to be analysed in view of constructing an index, we are confident that a model is possible.

Finally, we do not know whether the complexity of measuring "well-being", in the broad sense of measuring the performance of politics towards achieving the best of all possible worlds, will ever allow the calculation of a *politically accepted* overall wellbeing index. We are sure, though, that the mere *attempt* to construct such a holistic measure will teach us valuable lessons on how to find the best possible compromise between feasibility, scientific credibility and political acceptance.

---

1 http://www.encyclopedia.com/doc/1G1-17508781.html; note that Kuznets used the term "welfare", while The Economist refers to "well-being"; according to the Merriam-Webster Online Dictionary, welfare is defined as both "the state of doing well especially in respect to good fortune, happiness, *well-being*, or prosperity" and "aid in the form of money or necessities for those in need". The latter usage is prevalent when we speak of "welfare state". Since our intention is to capture well-being, not welfare in its narrow economic sense, we will use the term "well-being" throughout this article

2 See Rifkin 2004, p. 70 and following, for a discussion. Rifkin, J., 2004, The European Dream, Tarcher-Penguin, New York.

3 The latter include the JRC-CRELL/OECD Workshop on Measuring Well-being and Societal Progress (Univ. Cattolica, Milano 19-21 June 2006, http://crell.jrc.ec.europa.eu/Well-being.htm), the June 2007 Istanbul World Forum on measuring societal progress (http://www.oecd.org/oecdworldforum), and the April 2007 Rome Conference on subjective well-being (www.oecd.org/oecdworldforum/happiness)

4 Daly, H.E., Cobb, J.J. (1989) For the common good: redirecting the economy toward community, the environment and a sustainable future. Beacon Press, Boston

5 Pearce, D.W., Atkinson, G.D. (1993) Capital theory and the measurement of sustainable development: an indicator of "weak" sustainability. Ecological Economics, vol. 8, pp. 103-108

6 Yusuf, J.A., S. El Serafy, and E. Lutz, (1989) Environmental accounting for sustainable development. A UNEP World Bank Symposium, Washington D.C.

7 Vitousek, P., Ehrlich, P., Ehrlich, A. and Matson, P. (1986) Human appropriation of the products of photosynthesis. Bioscience, 34(6): 368-373

8 Wackernagel, M., Rees, W. E. (1995) Our Ecological Footprint: Reducing human impact on the earth. Gabriola Island, BC and Philadelphia, PA: New Society Publishers

9 Schmidt-Bleek, F. (1994) Wieviel Umwelt braucht der Mensch? MIPS, Das Mass für ökologisches Wirtschaften. Birkhäuser, Berlin

10 As a second common characteristic, we might add that the subcomponents included in these aggregate indices are *ad hoc*. For example, no clear justification is given of why *diet* enters in the computation of the Ecological Footprint while the generation of *waste* does not. However, in the absence of consensual rules on what must be included in such indices, we prefer to limit ourselves to the observation that the choice of a "magic" unit, whether

euros, dollars, hectares or tonnes, tends to silently but inevitably exclude variables that are politically important but refuse to be measured in this particular unit.

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

12 Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffmann A., Giovannini E., (2005) Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD Statistics working paper series

13 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

14 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

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

16 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

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

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

19 Saltelli, A. Tarantola, S., Campolongo, F. and Ratto, M. (2004) Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. John Wiley & Sons publishers, New York.

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

21 “Although Russia came top out of all countries analyzed in the sustainability of public finances and current account balance 2006 categories, it came last in the two environmental indicators”, Germany as a Business Location, http://www.allianz.com/en/allianz_group/press_center/news/studies/archive/downloads/study_location_ranking.pdf

22 Authors with a background in biology and with an interest in population growth, such as Paul Ehrlich and his collaborators, have over the years become aware of the shortcomings of the idea of carrying capacity as applied to humans. This is why they proposed the formulation I = PAT, I standing for human impact on the environment, P for human population, A for affluence, and T for technology. The definition of carrying capacity is irrelevant for humans for several reasons. First, the human ability to establish large differences in the exosomatic use of energy and materials points to a crucial question: at which level of consumption should the maximum population be established? Second, human technologies play a paramount role. To give an example transport is essential for determining urban carrying capacity because it influences the number of people which can enjoy a reasonable quality of urban life. Third, the territories occupied by humans are not given. We compete for them with other species as well as with other humans, whereby the concept of territory is socially and politically constructed. There is yet another reason why the notion of carrying capacity is not directly applicable to humans in any particular territory: trade may in fact be interpreted as the appropriation of the carrying capacity of other territories.

23 Joan Martinez-Alier, Giuseppe Munda, John O’Neill, 1998, Weak comparability of values as a foundation for ecological economics, Ecological Economics, 26, 277–286

24 See Saltelli, A., 2007, Composite indicators between analysis and advocacy, Social indicators research 81(11), 65-77

25 Manly B., (1994), Multivariate statistical methods, Chapman & Hall, UK

26 Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika. 16, 297-334

27 Anderberg, M.R. (1973), Cluster Analysis for Applications, New York: Academic Press, Inc.

28 George A. Miller, Harvard University, The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information, first published in Psychological Review, 63, 81-97 (1956).

29 Most known composites, however, treat sub-indices/dimensions as if they were individual indicators, i.e. each sub-index receives equal weight when calculating the top level score, *irrespective* of the number of underlying indicators. This neutral aggregation rule is also being used by the JRC/IISD Dashboard of Sustainability software, which has been applied to numerous indicator sets including the Millennium Development Goals and the Yale/Columbia ESI.

30 Nicoletti G., S. Scarpetta and O. Boylaud, (2000), Summary indicators of product market regulation with an extension to employment protection legislation, Economics department working papers NO. 226, ECO/WKP(99)18. http://www.oecd.org/eco/eco

31 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

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

33 The promoters of the BOD believe that "a particular dimension is deemed to be important for a country if the country performs well in that dimension. For example, it is assumed that the policy-makers of a country that performs well on the field of inflation will probably attach a high weight to a low inflation rate" (The Macroeconomic Performance of Nations, Moesen & Cherchye, Leuven 1998). Applied to a CI with three variables: GDP per capita, CO_2 emissions and life expectancy, this logic would lead to the conclusion that, when comparing OECD and Sub-Saharan Africa countries, it seems that the Africans are happy to be poor, eager to protect the global climate, and want to die early.

34 Green W.H., (2000), Econometric Analysis, Prentice-Hall International

35 Moldan, B. and Billharz, S. (1997) Sustainability Indicators: Report of the Project on Indicators of Sustainable Development. SCOPE 58. Chichester and New York: John Wiley & Sons

36 Saaty, R.W. (1987) The analytic hierarchy process - what it is and how it is used. Mathematical Modelling, vol.9, 161-176.

37 Hair J.F., Anderson R.E., Tatham R.L., and Black W.C., (1995), Multivariate data analysis with readings, fourth ed. Prentice Hall, Englewood Cliffs, NJ

38 Green P.E., and Srinivasan V., (1978), Conjoint analysis in consumer research: issues and outlook. Journal of Consumer Research 5, 103-123.

39 See http://www.externe.info/projects.html; more in Martin O'Connor & Anton Steurer (2006), The AICCAN, the geGDP, and the Monetisation Frontier: a typology of 'environmentally adjusted' national sustainability indicators, Int. J. Sustainable Development, Vol. 9, No. 1, 2006

40 Markandya, A., Pearce, D. (1989) Blueprint for a Green Economy. Earthscan, London

41 Inter alia, it is practically impossible to identify long-term reference technologies, and in particular to calculate true societal costs of avoiding CO2 emissions. As shown in Weizsäcker & Jesinghaus, Ecological Tax Reform (ZED 1992 and http://esl.jrc.it/dc/etr/ecological_tax_reform.htm), the long-term habit of European nations to heavily tax road fuels has halved EU consumption as compared to U.S.A and Canada, apparently at *no* overall cost to society. So reducing CO2 emissions through taxation might well have a *negative* cost to society, at least in Northern America. The implications for calculating SNI are pretty odd.

42 World Bank, Changes In Wealth, Chapter 3. Recent Genuine Saving Estimates, Footnote 3: "Tol (2005) reviewed over 100 estimates of the marginal damage cost of carbon dioxide emissions. He found a large range of uncertainty: the median cost was found to be $14 per ton of carbon and the mean to be $93/tC. On balance the use of the Fankhauser (1994) estimate of $20/tC appears to be reasonable." (http://siteresources.worldbank.org/INTEEI/214578-1110886258964/20744881/Chapter3.pdf downloaded 22.08.2007); using the figures quoted in "Where Is the Wealth of Nations?" (http://www.wider.unu.edu/public-lectures/public-lectures-2006/WealthofNationsconferenceFINAL.pdf, Appendix 3: Genuine Savings Estimates by Country), it seems that the United States have "Genuine Savings" at 8.2% of GNI, including a subtraction of 0.3% for CO_2 damages. The 0.3% might reflect the weight that U.S. politicians assign to the Kyoto Protocol, but it is hardly compatible with European views on climate change. See also *Monetary valuation of CO_2 emissions and the sensitivity of assumptions*, http://esl.jrc.it/envind/idm/idm_e_12.htm#Fig10

43 John Talberth, Clifford Cobb, Noah Slattery, The Genuine Progress Indicator 2006: A Tool for Sustainable Development (http://www.rprogress.org/publications/2007/GPI%202006.pdf): "In one recent meta-analysis of 103 separate studies, Tol (2005) found a mean of $93 per metric tonne, or $89.57 in year 2000 dollars. Though hotly debated, we adopt this figure as a conservative starting point for incorporating carbon emissions damage into GPI accounts". In total, the GPI aggregates 25 monetised indicators to the overall Genuine Progress Indicator.

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

45 For a detailed analysis of over 100 CI's, see Romina Bandura (UNDP, Office of Development Studies), A Survey of Composite Indices Measuring Country Performance: 2006 Update, http://www.thenewpublicfinance.org/background/Measuring%20country%20performance_nov2006%20update.pdf

46 An example taken from the JRC/IISD tool "Dashboard of Sustainability" (http://esl.jrc.it/dc/): "Life expectancy" in the EU-27: Best=81 years (Sweden), worst=71 (Estonia); for a life expectancy of 80 years, Italy receives 900 of 1000 possible points: P=1000*(80-71)/(81-71). The Dashboard allows a variety of other weighting options.

47 Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffmann A., Giovannini E., (2005) Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD Statistics working paper series.

48 Since multiplication by zero yields zero, geometrical weighting fails for CI's with data gaps. This is unfortunately the case for most if not all indicators in the Eurostat SD database.

49 Munda G. and Nardo M. (2007) - Non-compensatory/Non-Linear composite indicators for ranking countries: a defensible setting, forthcoming in Applied Economics.

50 See Munda (2008)- Social multi-criteria evaluation for a sustainable economy. Springer, Heidelberg, New York.

51 Saisana M., Saltelli A., Tarantola S., 2005, Uncertainty and Sensitivity analysis techniques as tools for the quality assessment of composite indicators, Journal of the Royal Statistical Society A, vol.168,(2), 1-17

52 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. New York: John Wiley & Sons.

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