Abstract

Composite indicators and ranking systems aggregate multi-dimensional processes into simplified concepts that are often used for advocacy and policy consumption. Due to methodological issues doubts are often raised about the robustness of the composite indicators and the significance of the associated conclusions. The OECD/JRC Handbook on constructing composite indicators offers a menu of statistical methodologies and technical guidelines that can help constructors of composite indicators to improve the quality of their outputs.

1. Introduction

Composite indicators as tools to compare country performance are increasingly common. Simplistically a composite indicator synthesizes the information included in a selected set of indicators and variables. The output of a composite indicator is a set of scores indicating the relative performance of a country within a set of countries. A composite is thus a measure of similarity. On Google (taken here as a proxy of overall diffusion of the concept) and Scholar Google (taken as a proxy of academic interest) the words “composite indicator” in January 2009 obtained respectively 398,000 and 328,000 hits versus 35,500 and 992 in October 2005. The interest is also witnessed by the growing number of composites in the public arena –165 in 2006 according to Bandura (2006) with a 30% growth with respect to the previous year– and interest of the media for rankings and league tables.

But why do we need composite indicators? Acceptance of indicators and composite indicators is rather established when dealing with dimensions having the same measurement unit, usually money (think to widespread measures of income, inflation, unemployment, etc.). However, the debate is still open when dealing with complex concepts in social sciences like citizenship, wellbeing, cohesion, or learning (Table 1 schematizes the pros and cons in using composite indicators). The widespread use of indicators and composites in social sciences¹ has much less tradition than in other areas

---

¹ Social cohesion indicators (Council of Europe, Eurostat,OECD)
(like economics) and consensus on the definition of the framework is much more common for concepts like inflation than for abstract concepts like learning or citizenship. Furthermore, in social sciences the variables populating the framework have different measurement units (in the case of social cohesion, for example, economic, social and individual variables have to be considered together) which makes comparability more difficult.

The alternative to quantitative indicators is the use of qualitative information (cases description). This can be richer in detail but context dependent, thus hardly replicable in space and time. The prevailing view today is that evidence-based policy should temper, if not replace, opinion-based policy. As the saying goes, “without data, you’re just a person with an opinion”\(^2\), measuring what can be measured (and sometimes even the non measurable) is the current orientation\(^3\). We do not fully share this view but we acknowledge the increased effort to measure abstract concepts such as knowledge (see PISA or TIMMS surveys)\(^4\).

The main virtue of composite indicators is their usefulness for policy analysis in that they can summarise complex and sometimes elusive issues in wide ranging fields, e.g., environment, economy, society or technological development. Composites often seem easier to interpret than finding a common trend in many separate indicators and have proven useful in benchmarking country performance. However, composite indicators builders have to face a relevant degree of skepticism among statisticians, economists and other groups of users. This skepticism is partially due to the lack of transparency of some existing indicators, especially as far as methodologies and basic data are concerned. Composites can send misleading policy messages if they are poorly constructed or misinterpreted. Their "big picture" results may invite users (especially policy makers) to draw simplistic conclusions. On the other hand, a sound process of construction cannot remedy to an inadequate framework or to poor quality data. Thus, the creation of a composite indicator requires a balance between different aspects, all equally important in defining the quality and finally the usefulness of the composite.

Another common misconception that needs clarification is that a league table, which is the result of a composite indicator, does not have universal legitimacy; scores are only valid within a given set of hypothesis thus the use of league tables, albeit appealing in attracting public attention, has little theoretical ground. Rather, composites should give a set of reference points towards which to benchmark the performance of a single country. Composite indicators should be seen as a starting point for initiating discussion and attracting public interest. Their relevance should be gauged with respect to constituencies affected by the composite indicator.

---

\(^2\) More precisely: without traceable data, you’re a person with a questionable opinion;

\(^3\) For example the European Commission claims the need for evidence policy in a number of documents starting from the white paper on governance (COM 2001 (428)). See also Boaz and Nutley (2003), Gary and Pring (2004).

\(^4\) See PISA (http://www.pisa.oecd.org); TIMMS (http://timss.bc.edu/); The European Commission (DG EAC) has foreseen for the next years the pilot testing of learning to learn features.
Table 1. Pros and cons in the use of composite indicators (rearranged from Saisana and Tarantola, 2002)

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Can summarise complex or multi-dimensional issues in view of supporting decision-makers.</td>
<td>• May send misleading policy messages if they are poorly constructed or misinterpreted.</td>
</tr>
<tr>
<td>• Easier to interpret than trying to find a trend in many separate indicators.</td>
<td>• May invite simplistic policy conclusions.</td>
</tr>
<tr>
<td>• Facilitate the task of ranking countries on complex issues in a benchmarking exercise.</td>
<td>• May be misused, e.g., to support a desired policy, if the construction process is not transparent and lacks sound statistical or conceptual principles.</td>
</tr>
<tr>
<td>• Can assess progress of countries over time on complex issues.</td>
<td>• The selection of indicators and weights could be the target of political challenge.</td>
</tr>
<tr>
<td>• Reduce the size of a set of indicators or include more information within the existing size limit.</td>
<td>• May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action.</td>
</tr>
<tr>
<td>• Place issues of country performance and progress at the centre of the policy arena.</td>
<td>• May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored.</td>
</tr>
<tr>
<td>• Facilitate communication with general public (i.e. citizens, media, etc.) and promote accountability.</td>
<td></td>
</tr>
<tr>
<td>• Constructing/underpinning narratives for lay or literate audiences</td>
<td></td>
</tr>
<tr>
<td>• Comparing effectively complex dimensions with one another</td>
<td></td>
</tr>
</tbody>
</table>

This paper offers an overview of the various steps in the construction of a composite indicator and underlines the main issues and problems a practitioner would encounter. It also provides hints on the application of composite indicators in the domain of social sciences.

Before starting the analysis of the various steps in constructing a composite indicator we clarify some basic definitions (see Munda and Nardo, 2007).

**Dimension:** is the highest hierarchical level of analysis and indicates the scope of objectives, individual indicators and variables. For example, a composite indicator on sustainability can include economic, social and environmental dimensions; a composite indicator on active citizenship can include participation to elections and to voluntary associations, but also values and attitudes of the citizens. The development of attitudinal indicators requires economic, social and psychological dimensions.

**Objective:** an objective indicates the direction of change desired. For example, within the economic dimension GDP has to be maximised; within the social dimension social exclusion has to be minimised; within the environmental dimension CO2 emissions have to be minimised. This is not always obvious. Take for example international mobility of researchers: it may need to be minimized when the hierarchical level is the country and the scope of the analysis is to measure brain drain, or to be maximized when the
Hierarchical level is constituted by the OECD countries and the analyst wants to measure peer learning.

**Individual indicator**: it is the basis for evaluation in relation to a given objective (any objective may imply a number of different individual indicators). It is a function that associates each single country with a variable indicating its desirability according to expected consequences related to the same objective. For example, GDP, saving rate and inflation rate inside the objective “growth maximization”.

**Variable**: is a constructed measure stemming from a process that represents, at a given point in space and time, a shared perception of a real-world state of affairs consistent with an individual indicator. To give an example, in comparing two countries along the economic dimension, one objective can be “maximisation of economic growth”; the individual indicator might be R&D performance, the indicator score or variable can be “number of patents per million of inhabitants”. Another example: an objective connected with the social dimension can be “maximisation of the residential attractiveness”. A possible individual indicator is then “residential density”. The variable providing the individual indicator score might be the ratio persons per hectare.

A **composite indicator** or synthetic index is an aggregate of all dimensions, objectives, individual indicators and variables used. This implies that what formally defines a composite indicator is the set of properties underlying its aggregation convention.

### 2. Steps in the construction of a composite indicator

As explained in the OECD/JRC Handbook on constructing composite indicators (2008), there exists an “ideal sequence” of steps to construct a composite indicator. Each step is important, and the coherence of the whole process is equally important: choices made in one step can have important implications for other steps.

1. **Theoretical framework**

The composite indicators’ controversy can perhaps be put into context if one considers that composite indicators are models, in the mathematical sense of the term. Models are inspired from systems (natural, biological, social) that one wishes to understand. The biologist Robert Rosen (1991) noted that while a causality entailment structure defines the natural system, and a formal causality system entails the formal system, no rule of encoding the formal system given the real system, i.e. to move from perceived reality to model, was ever agreed.

The formalization of the system generates an image, the theoretical framework, that is valid only within a given information space. As result, the model of the system will reflect not only (some of) the characteristics of the real system but also the choices made...
by the scientists on how to observe the reality. When building a model to describe a real-world phenomenon, formal coherence is a necessary property, yet not sufficient. The model in fact should fit objectives and intentions of the user, i.e. it must be the most appropriate tool for expressing the set of objectives that motivated the whole exercise. The choice of which individual indicators to use, how those are divided into classes, whether a normalization method has to be used (and which one), the choice of the weighting method, and how information is aggregated, all these features stem from a certain perspective on the issue to be modeled. Reflexivity is thus an essential feature of a model since “the observer and the observation are not separated […] the way human kind approaches the problem is part of the problem itself.” (Gough at al., 1998).

No matter how subjective and imprecise the theoretical framework is, it implies the recognition of the multidimensional nature of the phenomenon to be measured and the effort of specifying the single aspects and their interrelation. Most of the issues described with a composite indicator are complex problems, think to concepts like welfare, quality of education, or sustainability. Complexity is reflected by the multi-dimensionality and multi-scale representation of the issue. If we accept a definition of the theoretical framework requiring the integration of a broad set of (probably conflicting) points of view and the use of non-equivalent representative tools then the problem becomes to reduce the complexity in a measurable form. In other terms non-measurable issues like sustainability need to be replaced by intermediate objectives whose achievement can be observed and measured. The reduction into parts has limits when crucial properties of the entire system are lost: often the individual pieces of a puzzle hide the whole picture.

As suggested by Box (1979): ‘all models are wrong, some are useful’. The quality of a composite indicator is thus in its fitness or function to purpose. This is recognized by Sen (1989), Nobel prize winner in 1998, who was initially opposed to composite indicators but was eventually seduced by their ability to put into practice his concept of ‘Capabilities’ (the range of things that a person could do and be in her life) in the UN Human Development Index

In practice, a framework should clearly define the phenomenon to be measured and its sub-components and select individual indicators (and weights) that reflect their relative importance and the dimensions of the overall composite. Ideally, this process would be based on what is desirable to measure and not which indicators are available. And the transparency of the whole exercise is essential in constructing credible indicators.

### 2. Variable selection

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.

---

5 This Index is defined as a measure of the process of expanding people’s capabilities (or choices) to function. In this case, composite indicators’ use for advocacy is what makes them valuable.
While the choice of indicators must be guided by the theoretical framework, the data selection process can be quite subjective as there may be no single definitive set of indicators. The lack of relevant data also limits the constructor’s ability to build sound composite indicators. Given a scarcity of internationally comparable quantitative (hard) data, composite indicators often include qualitative (soft) data from surveys or policy reviews. However the use of soft data entails the risk of introducing significant measurement errors in the overall composite scores. 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, is often required. Finally, one has to make sure that the type of the selected variables – input, output or process indicators – match the definition of the composite indicator.

3. Imputation of missing data

Often data sets are not complete. Some countries or some years could lack data on a relevant indicator. Imputation of missing data is the art of filling empty spaces in a data matrix. In the words of Dempster and 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.”

In general there are three methods for dealing with missing data: i) case deletion, ii) single imputation or iii) multiple imputation. The first one, also called complete case analysis, simply omits the missing records from the analysis. However, this approach ignores possible systematic differences between complete and incomplete samples and may produce biased estimates. Furthermore, standard errors will in general 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 and Rubin, 2002).

The uncertainty in the imputed data should be reflected by variance estimates. This allows taking into account the effects of imputation in the course of the analysis. 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, or whether extreme values influence the whole exercise. More comprehensive surveys can be found in Little and Rubin (2002), Little (1997) and Little and Schenker (1994).

4. Multivariate analysis

Unfortunately, individual indicators are sometimes selected in an arbitrary manner with little attention paid to the interrelationships between them. This can lead to indices which overwhelm, confuse and mislead both decision-makers and the general public. Some
analysts characterise this environment as ‘indicator rich but information poor’. The underlying nature and properties of the data need to be carefully analysed before embarking on 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 of the dataset: 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) and Cronbach Alpha (Cronbach, 1951) 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. In this type of analysis the most similar countries are grouped and studied separately. Cluster analysis (Anderberg, 1973) is a useful 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 can also be used to impute missing data (Step 3 above) by using the available information on the countries that belong to the same cluster with the country in question.

5. Normalisation

Given that composite indicators are a way of summarizing complex phenomena, most of the times the individual variables forming the composites have different measurement units. For example GDP can be expressed in Euro, unemployment in number of persons, health in number of diseases, survey data in high, medium, low, or important, not important. In order to avoid summing up apples and oranges then one need to normalise the variables in order to make them comparable.

There are many possible normalization methods and practitioners should take into account data properties, as well as the objectives of the composite indicator in choosing one method rather than another. One has to be aware that different normalisation methods will yield different results in composite scores. Therefore the robustness of the final rankings (scores) should be tested by employing different normalization techniques. Normalization techniques are summarized and discussed in Nardo et al. (2005).
6. Weighting

Central to the construction of a composite index is the need to combine in a meaningful way different dimensions measured on different scales. This implies a decision on which weighting model will be used and which procedure will be applied to aggregate the information. Weights should ideally be selected according to a theoretical framework that is established or at least clearly stated. Weighting implies a “subjective” evaluation, which is particularly delicate in case of complex, interrelated and multidimensional phenomena. The menu of weighting methods is rather large and increasing with the creativity of the practitioners. Ideally, weights should reflect the contribution of each indicator to the overall composite. Different weights may be assigned to component series in order to reflect their economic significance (collection costs, coverage, reliability and economic reason), statistical adequacy, cyclical conformity, speed of available data, etc.

Most composite indicators rely on equal weighting, i.e., all normalized 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 2002 Environmental Sustainability Index), or even stem from the lack of consensus on alternative solutions (as happened with the 2001 Summary Innovation Index). In any case, equal weighting does not mean "no weights", but implicitly suggests that the weights are equal. Furthermore, it may happen that - by combining variables with a high degree of correlation - one may introduce an element of double counting into the index. Moreover, if variables are grouped into dimensions and those 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.

Notice that there will almost always be some positive correlation between different measures of the same aggregate. Thus, a rule of thumb could be introduced to define a threshold beyond which the correlation is a symptom of double counting. On the other hand, relating correlation analysis to weighting could be dangerous when motivated by apparent redundancy. For example, in the e-business Readiness Index, 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 correlation analysis but also by analyzing the indicator versus the ensemble of the remaining indicators and the phenomenon being measured.

Weights may also reflect the 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 base-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). 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 the theoretical importance of the indicators. Weights, however, cannot be estimated if no correlation exists between indicators. Other statistical methods, such as the benefit of the doubt (BOD) is extremely parsimonious about weighting assumptions as it lets the data decide on the weights and is sensitive to national priorities (Melyn and Mosen, 1991; Cherchye et al., 2008). BOD employs linear programming tools to estimate an efficiency frontier to be used as a benchmark to measure the relative performance of countries. However, given that the weights are derived from the data, they are also subject to eventual data measurement errors.

Multiple regression models can handle a large number of indicators (see a standard textbook in econometrics like Green, 2000). This approach can be applied in cases where the model input are indicators related to various policy actions and the model output is the target. The regression model, thereafter, could quantify the relative effect of each policy action on the output. 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. 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. In any case, there is always the uncertainty that the relations captured by the regression model for a given range of inputs and output, may not be valid for different ranges.

Unobserved components is similar in spirit to the multiple regression models but it does not need an explicit value for the “dependent variable” as it treats it like another unknown variable to estimate. This advantage is counter-balanced by the inconvenient of the complexity in estimation and the computational expensiveness.

Alternatively, participatory methods that incorporate various stakeholders – experts, citizens and politicians – can be used to assign weights. This approach is feasible when there is a well-defined basis for a national policy (Munda, 2005a,b, 2007). For international comparisons, such references are often not available, or they deliver contradictory results. In the budget allocation approach, 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 (Moldan and Billharz, 1997). The budget allocation is optimal for a maximum of 10-12 indicators. If too many indicators are involved, this method can give serious cognitive stress to 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).
The analytic hierarchy process (AHP) (pair wise comparison of attributes, Saaty, 1987) 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 simply assigning arbitrarily given weights. The resulting weights are less sensitive to errors of judgement. However, since the AHP is based on comparisons of indicator pairs, it is applicable only to low numbers of indicators.

Conjoint analysis derives the worth of the single indicator 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 and utility function (Hair et al., 1995; Green and Srinivasan, 1978).

We recommend using different weighting methods and test their effect on final country scores (and ranking). Weights usually have an important impact on the value of the composite and on the resulting ranking especially whenever higher weight is assigned to individual indicators on which some countries excel or fail. This is why weighting models need to be made explicit and transparent. However the impact of weighting method should not over-emphasized. In our experience other factors have the same or a higher impact on final scores and rankings, like 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 (Jacobs et al., 2004).

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 ad implication of the used weighting system should be always made clear and tested for robustness. Soundness and transparency should guide the entire exercise.

7. Aggregation

Weighting is strongly related to how the information conveyed by the different dimensions is aggregated into a composite index. Different aggregation rules are possible. Individual indicators could be summed up, multiplied or aggregated using non linear techniques. Each technique implies different assumptions and has specific consequences.

Linear aggregation method is useful when all individual indicators have the same measurement unit and further ambiguities due to the scale effects have been neutralized (for details see Nardo et al., 2005), while geometric aggregations (in which indicators are multiplied and weights appear as exponents) are appropriate when non-comparable and strictly positive individual indicators are expressed in different ratio-scales. The absence
of synergy or conflict effects among the indicators is a necessary condition to admit either linear or geometric aggregations. Furthermore, linear aggregations reward base-indicators proportionally to the weights, while geometric aggregations reward more those countries with higher scores.

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). 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 scores on one indicator will need much higher score on the others to improve its situation if the aggregation of information is geometric. Thus in a benchmarking exercise, countries with low scores should prefer a linear rather than a geometric aggregation. On the other hand the marginal utility of an increase in the score would be much higher when the absolute value of the score is low. The resulting lesson is that a country should be more interested in increasing those sectors with the lowest score in order to have the highest chance to improve its position in the ranking if the aggregation is geometric. The opposite is true, i.e. a country has interest in specializing along its most effective dimensions, when the aggregation is linear.

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 environmental 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 sustainability, 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. Table 2 summarizes the relationship between aggregation and weighting methods indicating the combination of weighing and aggregation that guarantees no loss of mathematical properties.

As mentioned above, Munda and Nardo (2003, 2005) noticed that in linear aggregations of indicators representing cardinal information, weights, customarily conceived as ‘importance’ measures, act in practice as substitution rates, e.g. \( w_i/w_j \) is the ratio of substitution (or compensation) of indicator ‘i’ with indicator ‘j’. This may be perceived as an important limitation of a composite indicator. Imagine for example that an index of development is being created and that literacy is one of the input variables. One might argue that literacy should not be traded with GDP per capita. When one is not willing to accept this kind of trade offs, a non-compensatory multi-criteria approach can be applied.

The multi-criteria procedure (MCA) tries to resolve the conflict arising in countries comparisons as some indicators are in favor of one country while other indicators are in favor of another. This conflict can be treated at the light of a non-compensatory logic and taking into account the absence of preference independence within a discrete multi-criteria approach (Munda, 1995, 2007). The approach employs a mathematical formulation
(Condorcet-type of ranking procedure) to rank in a complete pre-order (i.e. without any incomparability relation) all the countries from the best to the worst after a pair-wise comparison of countries across the whole set of the available indicators. We offer here a ‘hand waiving’ description of the algorithm. Imagine we have three countries, A, B and C and we aim at ranking their overall performance according to N indicators. We build to this effect an ‘outranking matrix’ whose entries $e_{ij}$ tells us how much country ‘i’ does better than country ‘j’. $e_{ij}$ is in fact the sum of all weights of all indicators for which country ‘i’ does better than country ‘j’. $e_{ji}$ will likewise 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 scaled to unity. 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 multicriteria 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. With this approach no compensation occurs, to exemplify, a country that does marginally better on many indicators comes out better than a country that does a lot better on a few ones because it cannot compensate deficiencies in some dimensions with outstanding performances in others.

Note that the MCA method provides results in terms of country rankings, and not of an index, so we can only follow the country ranks through time, provided that no countries are added or removed from the set. Further, a better rank for a given country at two different time points may not indicate an overall improvement in that country, but a deterioration of other countries in the set.

Between the full compensability of additive aggregations and the non-compensability of the multicriteria method an intermediate solution is the geometric aggregation. For example if an hypothetical composite were formed by inequality, environmental degradation, GDP per capita and unemployment, two countries, one with values 21, 1, 1, 1; and the other with 6,6,6,6 would have equal composite if the aggregation was additive. Obviously the two countries would represent very different social conditions that would not be reflected in the composite. Using instead a geometric aggregation the first country of our simple example would have a much lower composite than the second (2.14 versus 6).

Multicriteria analysis, like any other method, has pros and cons. At least in its basic form this approach does not reward outliers, i.e. those countries having large advantages (disadvantages) in individual indicators since it keeps only the ordinal information. Another disadvantage is the computational expensiveness when the number of countries is high (the number of permutations to calculate grows exponentially).
Table 2. When to use what: compatibility between aggregation and weighting methods.

<table>
<thead>
<tr>
<th>Weighting methods</th>
<th>Aggregation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear⁵</td>
</tr>
<tr>
<td>Equal weighting</td>
<td>Yes</td>
</tr>
<tr>
<td>Principal Component/Factor analysis</td>
<td>Yes</td>
</tr>
<tr>
<td>Benefit of the doubt</td>
<td>Yes</td>
</tr>
<tr>
<td>Unobserved component models</td>
<td>Yes</td>
</tr>
<tr>
<td>Budget allocation</td>
<td>Yes</td>
</tr>
<tr>
<td>Analytic hierarchy process</td>
<td>Yes</td>
</tr>
<tr>
<td>Conjoint analysis</td>
<td>Yes</td>
</tr>
</tbody>
</table>

¹ Benefit of the doubt requires additive aggregation; similar arguments apply to unobserved components.
² At least with the multi-criteria methods requiring weights as importance coefficients.
³ With both linear and geometric aggregations weights need to be treated as trade-offs and not “importance” coefficients.

8. Robustness and sensitivity

The reader will recall from the introduction that composite indicators may send misleading, non-robust policy messages if they are poorly constructed or misinterpreted. The construction of composite indicators involves stages where judgment has to be made: selection of data, data quality, data treatment (e.g., imputation), data normalization, weighting method, weights, and aggregation method. All these sources of subjective judgment will affect both the ranking —changes are more likely among middle-of-the-road performers — and the message brought by the composite indicator in a way that deserves analysis and corroboration (Saisana and Saltelli, 2008; Saisana and Munda, 2008; Saisana 2008; Brand et al. 2007). Uncertainty analysis (UA) and sensitivity analysis (SA) can be instrumental in this respect (Saltelli et al., 2004, 2008). UA focuses on how the sources of uncertainty propagate through the structure of the composite indicator and affect the composite scores. SA studies how much each individual source of uncertainty contributes to the variance of a country’s composite indicator score or rank. The synergistic use of UA and SA has proven to be more powerful (Saisana et al., 2005; Tarantola et al., 2000) than the application of UA alone (Jamison and Sandbu, 2001).

The types of questions for which an answer is sought via the application of UA&SA are:

(a) Does the use of one construction strategy versus another in building the composite indicator provide actually a partial picture of the countries’ performance?
(b) Which countries have large uncertainty bounds in their rank (volatile countries)?
(c) What are the factors that affect the countries ranks?

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

3. Conclusions

Our society is changing so fast that we need to know as soon as possible when things go wrong. Without rapid alert signals, appropriate corrective action is impossible. This is where composite indicators could be used as yardstick. Whether or not one accepts composite indicators for the purpose of benchmarking performance, one might find itself, even unwillingly, exposed to a composite indicator published in the news.

In this paper, we presented composite indicators as multi-dimensional and multi-scale representations of complex phenomena and we briefly explored the main steps necessary for their sound construction. We emphasized the need for a coherent and viable theoretical framework and we underlined the main issues related to weighting and aggregation in relation to compensability.

Yet, formal coherence is only one aspect of the composite indicators controversy. The bottle-neck conclusion is that composite indicators should never be seen as a goal per se, regardless of their quality or underlying variables. Given they aim at depicting a multi-scale and multi dimensional phenomenon, composites should be seen as a starting point for initiating discussion and attracting public interest and concern. The way of presenting composite indicators is also not a trivial issue. Composite indicators must be able to communicate the picture to decision-makers and users quickly and accurately. Visual models of these composite indicators must be able to provide signals, in particular, warning signals that flag for decision-makers those areas requiring policy intervention.

References


