# Composite indicators: lights and shadows

#### Andrea.Saltelli@uib.no

Open Evidence Research, Universitat Oberta de Catalunya (UOC), Barcelona



Rheinisch-Westfälische Technische Hochschule Aachen, University of Aachen, November 2, 2020



CAETERIS ARE

**NEVER PARIBUS** 

#### www.andreasaltelli.eu

HOME ABOUT ME

PUBLICATIONS

S NEWS & VIDEOS

S RESOURCES

(i)

#### Tweets by @AndreaSaltelli



andrea saltelli @AndreaSaltelli

The consequences of quantification?

"the ambition was to create a leading scientific university that would rise in international rankings"

How France overcame the odds to build a research mega-campus nature.com/articles /d4158...



How France overcame the odds to build ...



Composite indicators What are they?





#### Eight factors further disaggregated into 44 sub-factors

|   |   | - |   |  |
|---|---|---|---|--|
| Ī | T |   | Î |  |

Constraints on Government Powers 1.1 Government powers are effectively limited by the legislature

- 1.2 Government powers are effectively limited by the judiciary
- 1.3 Government powers are effectively limited by independent auditing and review
- 1.4 Government officials are sanctioned for misconduct
- 1.5 Government powers are subject to non-governmental checks
- 1.6 Transition of power is subject to the law

One of the eight factors with its 6 sub factors … https://worldjusticeproject.org/sites/default/files/documents/WJP-ROLI-2019-Single%20Page%20View-Reduced\_0.pdf Ubiquity of composite indicators



# At times useful

#### Making the case for gerrymandering?



Nature June 2017 article on the mathematics of 'nailing' gerrymandering



"[US] ranked 55th of 158 nations — last among Western democracies — in a 2017 index of voting fairness (Electoral Integrity Project)"

Carrie Arnold, 2017, The mathematicians who want to save democracy, 200, NATURE, VOL 546, 8 JUNE 2017.

At times pathetically wrong

The Global Health Security Index, released 2019 to "spur measurable changes in national health security" in light of "high-consequence and globally catastrophic biological events"



Cameron, E.E. et al., Global Health Security Index. Building Collective Action and Accountability. Nuclear Threat Initiative & Johns Hopkins Center for Health Security (October 2019). Available at https://www.ghsindex.org/#l-section--map.

#### US and UK rank 1 and 2 respectively



M. Kaiser, A. T.-Y. Chen, and P. Gluckman, "Should policy makers trust composite indices? A commentary on the pitfalls of inappropriate indices for policy formation," arXiv.org, vol. 2008.13637, Aug. 2020.

How can a country ranked last in quality of health care, with a raging opioid pandemic, be rated first in 'preparedness'?

See also:

https://www.nybooks.com/articles/2020/10/22/best-health-care/

https://www.commonwealthfund.org/publications/newsletterarticle/us-ranks-last-among-seven-countries-health-systemperformance



Quality of composite indicators



# Specific elements of quality for composite indicators



| Step 1: Theoretical framework      |  |
|------------------------------------|--|
| Step 2: Data selection             |  |
| Step 3: Imputation of missing data |  |
| Step 4: Multivariate analysis      |  |
| Step 5: Normalisation              |  |
| Step 6: Weighting                  |  |
| Step 7: Aggregating indicators     |  |
| Step 8: Sensitivity analysis       |  |
| Step 9: Link to other measures     |  |
| Step 10: Visualisation             |  |

Source: https://composite-indicators.jrc.ec.europa.eu/?q=10-step-guide

History

# The first scoreboard?

Ian Hacking, 1990, The taming of chance, Cambridge University Press.







#### Ian Hacking

#### Statistics $\leftarrow \rightarrow$ nation state $\leftarrow \rightarrow$ Modernity

Leibnitz, 'philosophical godfather of Prussian official statistics'.

# His proposal to the Prince Frederik of Prussia, 1700



Gottfried Wilhelm Leibniz (1646–1716)



56 categories to 'measure the power of a state', the first scoreboard;

- number of marriageable girls,
- able bodied capable to carry arms,
- diseases,
- child mortality,
- • •
- number of Jews





#### Gottfried Wilhelm Leibniz (1646–1716)

Is a theory for composite indicators possible?

"The role [of statistical indicators] has increased significantly over the last two decades.

This reflects improvements in the level of **education** in the population, increases in the **complexity** of modern economies and the widespread use of **information technology**"



Jean-Paul Fitoussi, Amartya Sen, Joseph Stiglitz

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report, last accessed June 2017.



#### Elements for a comprehensive assessment of public indicators



Paul-Marie Boulanger

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC9216 2/lbna26921enn.pdf



Paul-Marie Boulanger 2014

Editor: Andrea Saltelli



Report EUR 26921 EN

CI as boundary objects, between analysis and advocacy, as:

- instruments of democratization of expertise;
- instruments of social discovery
- semiotic objects

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf A triadic conception of the sign as structure connecting three elements:

- the sign properly said (S);
- an object (O) and
- an "interpretant"(I)



Charles Sanders Peirce, the father of semiotics 1839-1914 "This monkey possess a sophisticated repertory of vocal signs for signaling the presence of a predator"



African vervet monkey (Cercopithecus aethiops)

# It can distinguish

- a terrestrial stalking one such as a leopard,
- an aerial raptor such as an eagle or
- a ground predator such as a snake



African vervet monkey (Cercopithecus aethiops)

## Sign **←→** Cry



#### $Object \leftarrow \rightarrow Predator$



### Interpretant **C** Behaviour



Composite indicators as instrumental to the creation of a new public, through a process of social discovery (J. Dewey)



John Dewey 1859-1952

Dewey, J., 1938. The Public and its Problems, Read Book Ltd. Edition, 2013. Why are 'social discoveries' needed?

Because there are 'publics' affected by transaction taking place somewhere else.

"[…] machine age has so enormously expanded, multiplied, intensified and complicated the scope of the indirect consequences […] that the resultant public cannot identify and distinguish itself"

Dewey, J., 1938. The Public and its Problems, Read Book Ltd. Edition, 2013.

Social facts – unlike physical facts, are only meaningful in a context of desired ends

From J. Dewey 'Social Science and Social Control' in John Dewey: The Later Works, 1925–1953: 1931– 1932, Vol. 6-ExLibrary,





Building a composite indicator can be seen as a process of social discovery for which a model of extended participation comes natural.



Frames and indicators are coproduced in the process which must be designed as to have a meaningful 'interpretant', or 'end-in-sight'

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf

Critique of composite indicators: the Fitoussi-Stiglitz-Sen report "a general criticism … frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components […]



Jean-Paul Fitoussi, Amartya Sen, Joseph Stiglitz

 $[\cdots]$  an aggregation procedure always means putting relative values on the items that are introduced in the index"

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report, last accessed June 2017.
"The problem is not that these weighting procedures are hidden, nontransparent or non-replicable – they are often very explicitly presented by the authors of the indices, and this is one of the strengths of this literature.



Jean-Paul Fitoussi, Amartya Sen, Joseph Stiglitz

# The problem is rather that their normative implications are seldom made explicit or justified"

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report, last accessed June 2017. Critique of composite indicators: Ravallion There are types two indices: those built on economic theory / monetary aggregates / shadow prices and all others (=mashup indices)



Martin Ravallion

+ existing measures of e.g. development or poverty (Human Development Index, HDI, the Multidimensional Poverty Index, MPI) are bad at coping with tradeoffs

Martin Ravallion, 2010, Mashup indices of development, Policy Research Working Paper 5432, The World Bank Development Research Group, http://documents.worldbank.org/curated/en/454791468329342000/pdf/WPS5432.pdf

To illustrate the distinction, consider two stylized examples of composite indices, both formed from the data on household assets and consumer durables found in the Demographic and Health Surveys (DHS). For index A the variables and their weights are set by the analyst, who has some concept of -economic welfare in mind, and thinks this is related to certain variables in the DHS, which are aggregated based on the analyst's judgments.  ${
m FOr}$ index B, the variables and weights are instead based on a regression model calibrated to another survey data set for which a comprehensive measure of consumption (though still containing measurement errors) could be derived. The model is calibrated to common variables in the

expenditure survey and the DHS, and the regression model is used to predict wealth in the DHS.

A is a mashup index, B is not.



Martin Ravallion



Can composite indicators do harm?

J. Z. Muller, The tyranny of metrics. Princeton University Press, 2018.

### Unintended consequences: a litany

- Goal displacement
- Short termism
- Diminishing utility
- Rule cascade
- Discouraging risk taking
- Discouraging innovation

- Rewarding luck
- Discouraging cooperation and common purpose
- Degrading work
- Time waste
- Loss of productivity



Weapons of math destruction: opaque, do harm, do scale



WEAPONS

Since composite indicators are here to stay, how can we make them defensible?

… or how can we
deconstruct them?

Tools for evidence appraisal such sensitivity analysis and sensitivity auditing can be useful to gauge (and possibly deconstruct or reinforce) these measures

#### Sensitivity analysis



## Series A Statistics in Society

Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators

M. Saisana, A. Saltelli, S. Tarantola

First published: 3 March 2005 Full publication history

DOI: 10.1111/j.1467-985X.2005.00350.x View/save citation

Cited by (CrossRef): 181 articles for updates

Explore this journal >



View issue TOC Volume 168, Issue 2 March 2005 Pages 307–323

| Assumption           | Alternatives                                       |
|----------------------|--|
| Number of indicators | <ul> <li>all six indicators included or</li> </ul> |
|                      | one-at-time excluded (6 options)                   |
| Weighting method     | <ul> <li>original set of weights,</li> </ul>       |
|                      | <ul> <li>factor analysis,</li> </ul>               |
|                      | <ul> <li>equal weighting,</li> </ul>               |
|                      | <ul> <li>data envelopment analysis</li> </ul>      |
| Aggregation rule     | <ul> <li>additive,</li> </ul>                      |
|                      | <ul> <li>multiplicative,</li> </ul>                |
|                      | <ul> <li>Borda multi-criterion</li> </ul>          |



#### Sensitivity analysis to compare volatility of ranking

Research Policy 40 (2011) 165-177



#### Rickety numbers: Volatility of university rankings and policy implications Michaela Saisana\*, Béatrice d'Hombres, Andrea Saltelli

Econometrics and Applied Statistics, Joint Research Centre, European Commission, Enrico Fermi 2749, 21027 Ispra, Italy

#### Sensitivity analysis to compare volatility of ranking



## An 'invasive' analysis as the developers' choices are questioned/varied







THES ranking 2008

Shanghai ARWU ranking 2008 Incidentally: these university rankings have also damaged the educational systems

« processus de Bologne (en 1999) + stratégie de Lisbonne (en 2000), → passage d'une logique de service public à une logique de marché, concurrentielle et gestionnaire »



One can test whether assigned weights correspond to real importance

Journal of the Royal Statistical Society



*J. R. Statist. Soc.* A (2013) **176**, *Part* 3, *pp.* 609–634

#### Ratings and rankings: voodoo or science?

Paolo Paruolo

University of Insubria, Varese, Italy

and Michaela Saisana and Andrea Saltelli

European Commission, Ispra, Italy

The linear aggregation paradox: weights are used as if they were importance coefficients while they are trade off coefficients An example. A dean wants to rank teachers based on 'hours of teaching' and 'number of publications ...

$$Y = 0.5X_1 + 0.5X_2$$

X<sub>1</sub>: hours of teaching X<sub>2</sub>: number of publications

... adding these two variables up she sees that teachers are practically ranked by publications alone



Dean's example:  $y=x_1+x_2$ . Estimated  $R_{ht}^2 = 0.0759$ ,  $R_{np}^2 = 0.826$  To obviate this the dean substitutes the model  $y=0.5x_1+0.5x_2$  with

 $y=0.7x_1+0.3x_2$ 

A professor comes by, looks at the last formula, and complains that publishing is disregarded in the department ...





Using a result from sensitivity analysis the scatterplots → numbers reflecting the importance of a variable



The straight line corresponds to  $R^2$ 

The variance of the moving average is a sensitivity measure





 $S_i$  is the expected fractional variance reduction that would be achieved on average if  $x_i$  could be fixed

Why?

$$S_i = \frac{V(E(Y|x_i))}{V(Y)}$$

$$V(E(Y|x_i)) + E(V(Y|x_i)) = V(Y)$$

Normalization paradox: weights are assigned as to add up to one. This is questionable.

## Given a simple CI $Y = w_1 x_1 + w_2 x_2$

If both  $x_1$  and  $x_2$  are standardized the importance of  $x_1$  is  $S_1 = \frac{w_1^2}{w_1^2 + w_2^2}$ and  $S_1 + S_2 = 1$  Thus the relative importance of  $x_1$ ,  $x_2$ 

is not 
$$\frac{w_1}{w_2}$$
 but  $\frac{w_1^2}{w_2^2}$ 

### ... and the absolute importance:

not 
$$\frac{w_1}{w_1 + w_2}$$
,  $\frac{w_2}{w_1 + w_2}$   
but  $\frac{w_1^2}{w_1^2 + w_2^2}$ ,  $\frac{w_2^2}{w_1^2 + w_2^2}$ 

|                     | $x_1$ | <i>x</i> <sub>2</sub> |   |
|---------------------|-------|-----------------------|---|
| Presumed importance | 10%   | 90%                   | $\frac{w_1}{w_1 + w_2}$ , $\frac{w_2}{w_1 + w_2}$           |
| Real<br>importance  | 1.2%  | 98.8%                 | $\frac{w_1^2}{w_1^2 + w_2^2} , \frac{w_2^2}{w_1^2 + w_2^2}$ |

This holds if we use our definition of importance (what expected fraction of the variance of Ywould be reduced on average if  $x_1$  could be fixed) but you can varify this empirically using

fixed) but you can verify this empirically using scatterplots

W1=0.1, W2=0.9

![](_page_68_Figure_1.jpeg)

ten times more important of 81 times?

![](_page_69_Figure_0.jpeg)

These values would indeed give  $S_1 = 0.1$  and  $S_2 = 0.9$ 

Comparing assigned weights versus measured importance for the 2009 and 2010 versions of the Human Development index

> series A Statistics

> > Society

Journal of the Royal Statistical Society

J. R. Statist. Soc. A (2013) 176, Part 3, pp. 609–634

#### Ratings and rankings: voodoo or science?

Paolo Paruolo

University of Insubria, Varese, Italy

and Michaela Saisana and Andrea Saltelli

European Commission, Ispra, Italy

![](_page_71_Figure_0.jpeg)

Life expectancy, 33% Adult literacy, 22% Enrollment education, 11% GDP per capita, 33%






Life expectancy, 33% Education, 33% GNI per capita, 33% ■ declared weight □ importance

HDI2009

#### HDI2010



HDI 2010 more coherent than HDI 2009

Original Research | Open Access | Published: 23 January 2020

## Quantitative Storytelling in the Making of a Composite Indicator

<u>Marta Kuc-Czarnecka</u> ⊠, <u>Samuele Lo Piano</u> & <u>Andrea Saltelli</u> <u>Social Indicators Research</u> **149**, 775–802(2020) | <u>Cite this article</u> **1548** Accesses | **8** Citations | **5** Altmetric | <u>Metrics</u>





#### Marta Kuc-Czarnecka and Samuele Lo Piano

### Studying the convergence of EU countries

→What if different stakeholders have different preferences?

→ Stakeholder #1 trusts the variables in the European Pillar of Social Rights, while #2,#3 and #4 want to see 'more'

| Stakeholder 1              | Stakeholder 2              | Stakeholder 3              | Stakeholder 4              |
|----------------------------|----------------------------|----------------------------|----------------------------|
| Access to labour<br>market | Access to labour<br>market | Access to labour<br>market | Access to labour<br>market |
| Fair working conditions    | Fair working conditions    | Fair working conditions    | Fair working conditions    |
| Social protection          | Social protection          | Social protection          | Social protection          |
|                            | Fairness                   | Health care                | Fairness                   |
|                            |                            |                            | Health care                |



year

One step further: deconstructing the implicit normative framing of an indicator

## Blurring lines:

"what qualities are specific to rankings, or indicators, or models, or algorithms?"



Elizabeth Popp Berman

E. Popp Berman and D. Hirschman, The Sociology of Quantification: Where Are We Now?, Contemp. Sociol., vol. in press, 2017.

## Producing numbers comes with obligations

Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-020-00557-0

OPEN

# From sociology of quantification to ethics of quantification

Andrea Saltelli<sup>®</sup> <sup>1⊠</sup> & Monica Di Fiore<sup>®</sup> <sup>2⊠</sup>





COMMENT · 24 JUNE 2020

#### Five ways to ensure that models serve society: a manifesto

Pandemic politics highlight how predictions need to be transparent and humble to invite insight, not blame.



**COMMENT** • 24 JUNE 2020



# Five ways to ensure that models serve society: a manifesto

Pandemic politics highlight how predictions need to be transparent and humble to invite insight, not blame.

Andrea Saltelli , Gabriele Bammer, Isabelle Bruno, Erica Charters, Monica Di Fiore, Emmanuel Didier, Wendy Nelson Espeland, John Kay, Samuele Lo Piano, Deborah Mayo, Roger Pielke Jr, Tommaso Portaluri, Theodore M. Porter, Arnald Puy, Ismael Rafols, Jerome R. Ravetz, Erik Reinert, Daniel Sarewitz, Philip B. Stark, Andrew Stirling, Jeroen van der Sluijs & Paolo Vineis



#### Mind the assumptions

Assess uncertainty and sensitivity

#### Mind the hubris

Complexity can be the enemy of relevance

#### Mind the framing

Match purpose and context

#### Mind the consequences

Quantification can backfire.

#### Mind the unknowns

Acknowledge ignorance

SUPPLEMENTARY INFORMATION

**1.** Additional information and references >260 references



Conclusions: CI – instructions for use

Awareness of the imperfections and non-neutrality of measures

Beware damage; mind the interpretant

Investigate properties and assumptions (uncertainty and sensitivity analysis, sensitivity auditing)

Use for social discovery; deliberative extended participation; quality as fitness for purpose (interpretant)

## Reading material

Becker, W. et al. (2017) 'Weights and Importance in Composite Indicators: Mind the Gap', in Roger Ghanem, David Higdon, H. O. (ed.) Handbook of Uncertainty Quantification. Springer. http://www.andreasaltelli.eu/file/repository/Full\_Copy\_CI\_Handbook\_2017.pdf

Kuc-Czarnecka, M., Lo Piano, S. and Saltelli, A. (2020) 'Quantitative storytelling in the making of a composite indicator', Social Indicators Research, **149**, 775–802. https://link.springer.com/article/10.1007/s11205-020-02276-0







## Alain Supiot



## A critique of normative uses of quantification

#### **Alain Supiot**

#### La Gouvernance par les nombres

Cours au Collège de France (2012-2014)



<u>https://www.college-de-</u> <u>france.fr/site/en-alain-</u> <u>supiot/Governance-by-Numbers-</u> <u>Introduction.htm</u>

FAYARD POIDS ET MESURES DU MONDE





Futures

Volume 116, February 2020, 102509



Essays

# Ethics of quantification or quantification of ethics?

Andrea Saltelli

https://doi.org/10.1016/j.futures.2019.102509

#### What do I make of your latinorum? Sensitivity auditing of mathematical modelling

Andrea Saltelli\* and Ângela Guimarães Pereira

European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, Via E. Fermi, 2749, 21027 Ispra (VA), Italy E-mail: andrea.saltelli@jrc.ec.europa.eu E-mail: angela.pereira@jrc.ec.europa.eu \*Corresponding author Jeroen P. Van der Sluijs

Copernicus Institute of Sustainable Development, Utrecht University (NL), Heidelberglaan 2, 3584 CS Utrecht, The Netherlands E-mail: j.p.vandersluijs@uu.nl

Silvio Funtowicz

Centre for the Study of the Sciences and the Humanities (SVT) Allegaten, University of Bergen (NO), 34 – Postboks 7805 5020 Bergen, Norway E-mail: silvio.funtowicz@svt.uib.no

#### nt, The Netherlands

Sensitivity auditing

rutures xxx (2017) xxx-xxx



Original research article

What is wrong with evidence based policy, and how can it be improved?

Andrea Saltelli<sup>a,b,c,\*</sup>, Mario Giampietro<sup>a,c,d</sup>

#### Too much is being read in the OECD-PISA data

IJCED 19,1

#### Do PISA data justify PISA-based education policy?

Luisa Araujo Department of Human Capital and Employment, European Commission Joint Research Centre Ispra Sector, Ispra, Italy Andrea Saltelli

University of Bergen, Bergen, Norway and Universitat Autonoma de Barcelona, Barcelona, Spain, and

Sylke V. Schnepf Competence Centre on Microeconomic Evaluation, European Commission Joint Research Centre Ispra Sector, Ispra, Italy



International Journal of Comparative Education and Development Vol. 19 No. 1, 2017 pp. 20-34 © Emerald Publishing Limited 2396-7404 DOI 10.1108/IJCED-12-2016-0023

20

Received 14 December 2016 Revised 17 February 2017 Accepted 24 February 2017



Academic rigour, journalistic flair

Arts + Culture Business + Economy Cities Education Environment + Energy FactCheck Health + Medicine Politics + Society Science + Technology



June 12, 2017 3.55pm AEST

#### RELEVANCE

In the context of composite indicators, relevance has to be evaluated considering the degree to which it meet current and potential needs of the users

[…] ensure that the right range of domains is covered in a balanced way

#### ACCURACY

The credibility of data products refers to confidence that users place in … the image of the data producer, i.e., the brand image …

[crucial] that the data are perceived to be produced professionally and that practices are transparent

(for example, data are not manipulated, nor their release timed in response to political pressure)

#### COHERENCE

ensure coherence over time and across countries
Coherence across countries implies that from country to country the data are based on common concepts, definitions, classifications and methodology, or that any differences can be justified