

International Ratings and Rankings: Cure or Disease?

Philippe C. Schmitter

European University Institute

Indexene: a remarkable new cure that has rapidly been taken by social scientists, especially by political scientists and even more especially by think-tankers and policy advocates, that allows them to solve the intrinsic problem of measuring and explaining complex, multi-dimensional phenomena by assembling numerical ingredients (usually already available from reputable suppliers) into a single compound indicator which will have the appearance of hard, quantitative science and command the attention and admiration of academics, decision-makers and the wider enlightened public.

Indexitis: a recently discovered and rapidly proliferating disease among social scientists, especially among political scientists and even more especially among think-tankers, that leads to the erroneous belief that assembling data (often of questionable origin, validity and accuracy) and simply adding them together (often with the presumption that they are all of equal importance) can produce an accurate and comparable indicator for complex, multi-dimensional phenomena (admittedly, more for policy and public consumption than for scientific inquiry).

Indexene has already been field-tested. The panelists assembled in Ranking the World. Grading States as a Tool for Global Governance have collectively declared it safe for social and political consumption, although with reservations.ⁱ They have expressed some doubts about its ingredients, its side-effects and its efficacy. The cure is supposed to be consistent, transparent, impartial, scientific, efficient, and, I would add, appealing to the policy-making consumer.ⁱⁱ Almost unanimously, they have noted that the components of indices are often assembled intuitively, i.e. without theoretical guidance and redundantly, i.e. count the same thing more than once because the components are so highly correlated. In other words, the remedy and the disease can be confused within the same treatment. Also, the dosage for different users can be the same, even if the individual scores are different, even diametrically opposed for each of the countries involved.ⁱⁱⁱ Some *indexene* producers simply add together the ingredients under the assumption that they are all of equal importance. Others weigh them according to some formula which may be intuitively convincing, but almost never explicitly justified. What is particularly alarming is that the

ingredients may be simultaneously and systematically present, but interact in unpredictable ways to produce radically different results. What is almost as damaging, however, is that indices may be labelled identically, but generate very different scores for the same country. The contribution by Seva Gunitsky on democracy indices applied to the former Soviet Republics is particularly illuminating in this regard.^{iv} Her graphs in Chapter 5 and in Appendix 2 show astonishingly different scores at any moment in time and over time – presumably based on different definitions of the disease rather than different clinical observations. The Transparency Perception Index (TPI) is often cited as one of the most successful when measured in terms of media citation or references by policy-makers. Yet it relies on the Delphi Method of extracting opinions from – of all people – foreign businessmen, when it is the *Hoi polloi* who are not likely to be known by these respondents and who are most likely to be affected by corruption in any given country.

Indexitis is a disease plaguing the recent flood of compound indicators of state performance and should be avoided. The core of the problem lies with the complexity of institutions and the policies they pursue, coupled with the recognition by virtually all social scientists that this has increased exponentially in recent decades. This already difficult analytical challenge has been made worse by the seemingly unlimited demand by existing policy-makers and aspiring policy influencers for simple ratings or rankings that they believe will help them make more enlightened policies or attract the attention of relevant publics. Social scientists are almost equally guilty since they are often under peer pressure to appear scientific and believe that this requires them to convert difficult and multi-dimensional qualities into simple, aggregative quantities and, of course, to manipulate them statistically.

Recognizing the effectiveness of blaming and shaming specific countries – especially when these are places whose behavior the index-maker does not approve of – is not the same thing as pretending that such data are scientifically valid and, hence, capable of providing a useful compound indicator of such complex phenomena as democracy, sovereignty, human rights, regime transformation, stateness, rule of law, quality of life *e così via*.

Institutional or policy complexity is not additive. It is rooted in the interaction of multiple components and subject to the contingencies imposed by the environments in which they are embedded. Even when (not often the case) the components chosen, standard scored, weighted and combined to form an index are informed by a plausible theory, this does not capture either the interaction effects or the environmental contingencies. Many of these indices are composed by simply summing their variables – as many as 24!^v It is entirely possible that two or more units could receive the same aggregate score while differing (even being diametrically opposed) on every item. Inversely, it is usually possible to produce more empirically reliable indicators, as judged by inter-coder agreement, by measuring the same thing several times or by convoluting cause and effect.

The obvious solution for scholars is not to stop inventing new indicators and gathering additional data, but to stop thoughtlessly aggregating these data into a single index. Granted that consumer appeal and public impact would suffer, but this could open the door to the systematic analysis of how the component items relate to each other and respond to different contextual conditions.

* * *

Indexitis can be cured, but it will not be easy since the prestige of so many social scientists and the clout of so many policy advocates are at stake. What is needed is not some massive antidote and certainly not an enema to purge it altogether. What is needed is a radical separation between two different cures: *Indexene₁* and *Indexene₂*. The former would be designed intuitively by activists in think-tanks, government agencies and non-governmental organizations and compiled for maximum public impact. The latter would be designed by scholars according to a careful examination of its internal ingredients and their interrelationships for the purpose of scientific inquiry. *Indexitis* emerges when the two are confounded – when advocates pretend that their *Indexene₁* scores and rankings are conceptually valid and empirically accurate and when the *Indexene₂* of scholars fails to reflect theoretical priorities and/or to respect internal consistency. The result of such a separation could a more healthy future for all concerned.

ⁱ Alexander Cooley & Jack Snyder (eds.), Ranking the World. Grading States as a Tool for Global Governance (Cambridge: Cambridge University Press, 2105). Actually the chapter by Nehal Bhuta on “Measuring Stateness, ranking political orders: Indices of state fragility and state failure,” pp. 85-111, comes close to advocating prohibition.

ⁱⁱ Op. cit., p. 178.

ⁱⁱⁱ This seems to increase monotonically with the number of items including in the index.

^{iv} “Lost in the Gray Zone. Competing Measures of democracy in the former Soviet republics,” op.cit., pp. 112-150.

^v This seems to be the case of the Globalization Index of the Economic Intelligence Unit – although the items are weighted according to an explicit formula which, however, is not justified theoretically, nor is the inclusion of specific items).