

POLICY DIRECTIONS

MEASURING POOR STATE PERFORMANCE



Francisco Gutiérrez, Andrea González, Diana Buitrago, Camila Lozano

HIGHLIGHTS:

- **The policy community is working with seriously flawed comparative measurements of state fragility, which at best are poor guides to making policy and at worst can lead to major distortions.** Existing criticisms of the quantitative ranking of states by their level of ‘fragility’ emphasise the difficulty of producing an understanding of state fragility that is reasonable, accurate and useful. Through various publications (see for example [Gutiérrez and González 2009](#); [Gutiérrez 2009](#)) the research team adds to this body of criticism, but – more constructively – argues that despite the difficulties associated with measuring state fragility, *it is an essential task*.
- **Our researchers have devised an alternative approach to measuring fragility that overcomes some of the important problems.** Because of the type of criticism that we have developed, it was possible for the IEPRI¹ team to overcome many of the problems of existing measurements. Instead of attacking the outcomes of the most common rankings of fragility, the focus of attention is placed upon the methodological shortcomings of those exercises, and in particular on their failure to produce credible aggregations. The team has *formally* demonstrated (that is, through systematic explanation) that the ranking of countries in existing indexes almost always depends on how index designers aggregate the values of different variables. As we explain below, most of these indexes have either not recognised this problem or employed non-credible methods to aggregate the data in their index. The IEPRI research team put in operational form, or created, aggregation procedures that address key issues that until now have remained unacknowledged.
- **The research team has built a new database to measure state performance.** Besides improving the aggregation tool, the team built a new database – covering Monopoly of violence-Administration-Territory (MAT). This is consistent with specific theories about the state and with the wealth of research produced by the CSRC, and does not collapse definitional traits with putative causes and consequences. For instance some indexes include in their *definition* of fragility, poor economic performance or a lack of democracy, when in fact the measurement exercise would call for these traits to be considered either as a cause or consequence of fragility. Collapsing definitions, putative consequences, and putative causes would be like including smoking (a cause) and familial crisis (a consequence) in the definition of cancer, along with the malign proliferation of cells. It prevents the proper isolation and analysis of phenomena, a prerequisite of sound policy formulation.
- **We have created new tools that deal with the incomplete and ambiguous nature of the data.** Typically, builders and managers of political indicators have to deal with incomplete, deteriorated, intrinsically ambiguous, and highly multidimensional data. The team created specific tools (fuzzy regressions and clustering, for example) that have allowed us to produce systematic data analyses while taking into account any limitations of the specific type of data that is being utilised.
- **New conclusions about state fragility are emerging.** Initial analysis stemming from the MAT database with the new tools is starting to yield interesting conclusions: Not all poor countries are fragile; different aspects of fragility have separate causes; and achieving state resilience may make progress in development difficult

The CSRC Policy Directions series highlights the analysis, evidence and key policy recommendations emerging from CSRC research, funded by the Department for International Development

www.crisisstates.com



¹ Instituto de Estudios Políticos y Relaciones Internacionales, Universidad Nacional de Colombia



MODERN STATES NEED TO MEASURE

Indexing, or classifying reality, is a necessary condition for the existence of a modern state; indexes not only describe reality but constitute it (Bouyssou et al 2000). In creating authoritative forms of counting and developing the criteria that pick out 'winners', indexes have become both measurement and incentive systems and thus can help understanding, as well as promoting change. Moreover, these indexes have enabled several actors – powerful states and transnational agencies in particular – to produce, follow and administer global public policies. There are specific and relevant aspects of such policies that can only be captured by aggregated measurements.

What can be termed a 'third wave' of indexes has emerged over the past twenty years, each of which is related to a technological revolution, the globalisation of capitalism and politics, the creation of global audiences, and the development of global policies that must be followed, evaluated and adjusted. These are no longer simply 'counts': these new indexes incorporate numerous variables, have a complex structure, use diverse sources of information, and quantify multi-layered and highly contested concepts – like state fragility.

In their present form, the majority of Poor State Performance Indices (PSPIs) are unsound. The reliability and validity of such high-profile PSPIs have been frequently challenged. Whilst not all criticisms are valid, or even relevant, some key points hold firm: many of the PSPIs are guilty of 'conceptual stretching' where lack of definition and too many variables produce arbitrary and 'noisy' results; aggregation (or weighting) of the variables is inadequately or incorrectly done, and in reality never discussed despite being the crucial step of index building after making the concept operational; and definitional inconsistency, or the assigning of ad-hoc cut-off points for data groupings, diminish the usefulness of results.

The strength of criticism against PSPIs is not surprising. Numerous factors make measuring state fragility difficult: the underlying concept is inherently complex, and numerous dimensions are thus required to capture it. Many potential biases may affect the variables measured. The variables are heterogeneous and can include, for example, press monitoring, expert judgement and non-representative polls, which are difficult to manipulate numerically. Data, where possible to collect, tends to be ambiguous, incomplete, and deteriorated. There is no obvious counting unit: what is a unit of democracy, and how can this be compared to a unit of state fragility? The lack of genuine hard data is frequently replaced by scales, which are suspect on many accounts (they do not represent true orderings). We show that – while many of these problems are idiosyncratic – there are objective limits to building highly multi-dimensional indicators with ambiguous and otherwise imperfect data. It will be impossible to arrive at a 'bullet proof' indicator. However, the alternative is nothing: no results, no data, no analysis. Whilst the aspiration of having bullet proof indicators, good for all contexts and situations, may be a case of excessive ambition, it is possible to arrive at appropriate solutions tailored for specific problems and contexts.

Developing and improving formal decision tools is important in a world where there are many settings, agencies, and officials related to global decision making. Increasingly, we have to deal

with phenomena that affect many countries, in different ways. Research on individual countries through in-depth case study will not allow researchers or decision makers to ignore the key issue of aggregation, this is, the combination of different criteria to produce a single evaluation that makes situations/countries comparable. For when we speak about policy and change we speak basically of aggregation, because policy and change are complex: different units of analysis will inevitably be affected in different ways. Furthermore, aggregation tools hold the potential to advance understanding of the way in which a given phenomenon relates to other factors across countries, continents, and levels of development. Thus quantitative, and more generally formal analytical and decision tools, are not substitutes but complements to good, contextually dense, qualitative research.

CRITICISING THE CRITICS

The CSRC is well placed to mount a vehement and comprehensive critique of PSPIs, underlining the strength of the already established anti-ranking scale discourse. First and foremost, the CSRC has hosted an active triangulation and dialogue between qualitative and quantitative research, and a conceptualisation that can be employed by both approaches. Second, as a research undertaking engaged in constant dialogue with global decision makers but independent of them, we have been able to go beyond immediate policy concerns while addressing policy-relevant issues.

Simply criticising PSPIs gets us nowhere. It is necessary to look past these flaws, because indicators, despite all of their limitations, are tools directly related to the concept of statehood and because of the need to produce and evaluate global decisions. Thus, the research team has gone two steps further: first mounting a defence of existing PSPIs against some of the more untenable or irrelevant criticisms, and second suggesting a new and different measurement mechanism that addresses some of the most significant problems with existing PSPIs while acknowledging the fact that other problems will inevitably remain.

1. Simplifying reality

Indexes are supposed to be extreme simplifications of reality. The demand that they describe the complexity of the context is incorrect. That is the job of qualitative research. Indexes are powerful precisely because they are simple: this makes them tractable and enables abstract but systematic comparisons. Ordering of states, for example, is useful if – and only if – it is appreciated that this ranking is not exact or absolute.

Indexes do not reflect *all* reality, but again, they are not supposed to. To demand that indexes should incorporate the numerous variables that constitute reality results in confusion and invalidity. Measuring growth and inequality at once negates the possibility of seeing how they covariate, or how they interact.

2. Comparing realities

Contrary to popular criticism, one of the functions of indexes is precisely in order to enable a comparison of apples and oranges. This may produce more questions than answers at times, but these interrogations are important and hold insight in themselves. Indexes do not tell a whole story, but nor are they supposed to do so. A comparison based on defensible categorisation and quality of data is legitimate and interesting: defensible does

not mean exact. For example, whilst measurement of GDP in DR Congo and Bolivia might not produce entirely reliable and accurate data, it is still useful to know that DR Congo's GDP is lower than Bolivia's.

On the other hand, at least two fundamental problems remain. On the one hand, PSPI and other political databases are specific from the point of view of the type of data that they administer. Their databases 'live' in a much more multi-dimensional and complex space than others, like the datasets measuring growth. They try to operationalise concepts that are ambiguous and have no properly defined borders: when, for example, does an 'adequate' provision of public services start or end? Almost all PSP variables therefore have crucial missing information, especially where it is most needed (such as information related to countries in deep conflict or suffering major disruption). Many indexes rely on expert assessments of general social processes, without clearly noting that they contain specific variables that can increase or decrease separately. For example, a variable about 'repression' or 'violence against civilians' is not really ordinal because it includes counts that do not necessarily increase simultaneously. Some forms of violence may go up, while others go down. In sum, PSPIs have to deal with deteriorated, incomplete, and ambiguous data.

Second, all aggregation procedures can run into problems; this is particularly true when used in highly dimensional databases that are intended to operationalise complex concepts. For example, no reasonable aggregation procedure devised to produce a state performance index can be fully compensatory: that is, it cannot assign a simple trade-off between one person who was tortured and the sound delivery of an immunisation programme. To do so is to embed in the index assumptions of equivalencies that are patently absurd and indefensible.

Until now, all indexes have used fully compensatory aggregation rules, and resulting rankings are thus deeply flawed.

The research team advanced from this critical stand in four ways. First, we developed the *formal* understanding of the implications of high multidimensionality for index creation. We showed that the number of variables grows very fast in possible aggregation procedures, and thus rankings. With a big number of variables, there are many different ways of aggregating the data. A haphazard choice of any of these numerous possibilities will determine artificially a potentially very important part of the ranking. Second, after demonstrating this, we developed possible aggregation procedures, which performed well according to a set of simple, clear, and explicit assumptions. Importantly, none of them is fully compensatory, and thus we avoid the problems mentioned above. Third, some of these aggregation procedures produced not numbers, but other kinds of objects (like numerical intervals). Previous literature had noted the possible representational importance of such objects, but until now had not devise a way of creating them, and utilising them, in a systematic fashion. The team built the tools to do so. The importance of such representations is that they explicitly take on board the ambiguity of both definitions and data, shedding at least part of the spurious precision found in many of the typical quantitative exercises that use political indexes.

THE MONOPOLY ADMINISTRATION-TERRITORY (MAT) DATABASE

The fourth step was the building of a new database itself. The departure point was very simple: use clear, established concepts, and try to trim down dimensionality and avoidable ambiguity.

The concept of statehood that guides the construction of the MAT database draws heavily on the Crisis States Programme (Phase 2) approach and results (see especially [Di John 2008](#) and [Putzel 2006](#)), though it is not a literal transcription of them. For CSRC Phase 2, as well as for the bulk of social history and political economy analyses of the state, the core of statehood is composed of three elements: monopoly of violence, bureaucracy, and the territorial reach of the state. According to Tilly (1975; 1990), statehood is defined by a continuum that goes from oligopoly to the monopoly of violence and coercion. Bureaucracy is the signature of modern states, and there is also a broad consensus in the social sciences that it is both a key characteristic of statehood in its own right and a proxy of the capacities of the state (see for example Weber 1922, and Mann 1984). Indeed, what matters most here is not the size of the bureaucracy, though at the limits it should indeed be decisive,² but its efficacy. Finally, the ability to control a contiguous territory and to operate on it is at the very heart of the definition of modern sovereignty (see for example North, Wallis and Weingast 2009; Blanton and Fargher 2008; Jackson 1990).

STATE FRAGILITY, DEMOCRACY AND DEVELOPMENT

The use of the tools described above has helped the team to develop and specify some *preliminary* conclusions. Some of the main ones are as follows:

- a. States cluster around differential behaviour. While some are good or bad performers in all dimensions, the situation of good performance in dimension A and poor performance in dimension B is much more typical. This underlines an existing and important CSRC conclusion – that states perform differentially across their functions – but also will permit the full formal characterisation of the phenomenon; that is, allow us to investigate it systematically using mathematical tools.
- b. The correlates of performance in each dimension of statehood are also different.
- c. Competitive democracy, as captured by Polity, has limited explanatory power. Democracy, however, gets explanatory punch when the proxy for 'non-repressiveness' is included.
- d. Savings as a percentage of GDP seem to be strongly and negatively correlated with fragility.
- e. Having a neighbour that has lost the monopoly of violence is an important source of fragility.
- f. There is a strong variance at the bottom of the development ladder – a genuine 'multilevel effect'. While at the top of the ladder only strong and democratic states are found, at the

² In other words, if there is only one employer, the state, this is a significant datum. On the other hand, if the state almost has no employees, that is a significant datum as well.

bottom there are many types of states (fragile and non-fragile, etc.). Low development and fragility are distinct phenomena. Analysis and policy have to identify the correlates of the different levels of fragility at the bottom.

Not all poor countries are fragile. If development (measured by GDP), and state fragility (measured by an aggregated index coming from the MAT variables) are correlated then a huge amount of unexplained variance remains and, furthermore, this variance is especially large at the bottom. This means that there are many different types of states at the lowest levels of development: some quite resilient, others quite fragile. Understanding that these different types exist and that different responses will thus be necessary should be the central focus of the policy community in relation to fragile states

Different aspects of fragility have separate causes.

Each component of state performance seems to be associated with a different set of variables. This confirms, potentially develops and engages with two of the key findings of the CSRC's qualitative research.

Achieving state resilience may make progress in development difficult. Analysis of the relationship between development and state-making under the CSRC programme has required significant engagement with poor state performance indexes (PSPIs). One of the main conclusions of the CSRC's 'Development as State-Making' research is that resilient states may be a result of a trade off at any given time between securing peace and promoting positive programmes for economic growth and welfare.

Some of these results are quite robust, and persist across different models, types of regressions, and periods. We believe this approach to measuring state performance needs to inform the decisions taken by the policy and practice communities.

REFERENCES

- Blanton, R., and Fargher, L. (2008). *Collective Action in the Formation of Premodern State*. New York: Springer.
- Bouyssou, D., Marchant, T., Pirlot, M., Perny, P., Tsoukias, A., and Vincke, P. (2000). *Evaluation and Decision Models. A Critical Perspective*. Massachusetts: Kluwer Academic Publisher.
- Di John, J. (2008). Conceptualising the Causes and Consequences of Failed States: A Critical Review of the Literature. *Working Paper 25(2)*, Crisis States Research Centre.
- Gutiérrez, F. (2009). The Quandaries of Coding and Ranking: evaluating poor state performance indexes. *Working Paper 58(2)*, Crisis States Research Centre.
- Gutierrez, F. and González, A. (2009). Force and Ambiguity: evaluating sources for cross-national research - the case of military interventions. *Working Paper 50(2)*, Crisis States Research Centre.
- Jackson, R. (1990). *Quasi-States: Sovereignty, International Relations, and the Third World*. New York: Cambridge University Press.
- Mann, M. (1984). The Autonomous Power of the State: Its Origins, Mechanisms and Results. *Archives Européennes de Sociologie*, 25, 185-213.
- North, D., Wallis, J., & Weingast, B. (2009). *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. New York: Cambridge University Press.
- Putzel, J. (2006). War, state collapse and reconstruction. *Working Paper 1(2)*, Crisis States Research Centre.
- Tilly, C. (Ed.) (1975). *The Formation of National States in Western Europe*, Princeton, Princeton University Press.
- Tilly, C. (1990). *Coercion, Capital, and European States, AD 990-1990*. Cambridge: Cambridge, Mass., USA: B. Blackwell.
- Weber, M. (1922). *Economy and Society*. México: Fondo de Cultura Económica.

All CSRC papers referenced in this brief are accessible by hyperlink or on www.crisisstates.com/publications/publications.htm



Crisis States Research Centre
LSE
Houghton Street
London WC2A 2AE
e: csp@lse.ac.uk
www.crisisstates.com



Printed on recycled paper

THE CRISIS STATES RESEARCH CENTRE

The Crisis States Research Centre (CSRC) is a leading centre of interdisciplinary research into processes of war, state collapse and reconstruction in fragile states. By identifying the ways in which war and conflict affect the future possibilities for state building, by distilling the lessons learnt from past experiences of state reconstruction and by analysing the impact of key international interventions, Centre research seeks to build academic knowledge, contribute to the development of theory, and inform current and future policy making.

The Centre is based within the Development Studies Institute (DESTIN) of the London School of Economics and Political Science and is funded by a grant from the UK Department for International Development.

Readers are encouraged to quote this publication but CSRC requests acknowledgement for purposes of copyright. Views expressed within do not necessarily reflect those of LSE or UKaid.