Response to "What Do the Worldwide Governance Indicators Measure?"

Daniel Kaufmann , Aart Kraay, and Massimo Mastruzzi[1]

July 14, 2009

Forthcoming, European Journal of Development Research

Abstract: Thomas (2009) dismisses the Worldwide Governance Indicators (WGI) as an "elaborate and unsupported hypothesis" because of the failure to demonstrate the "construct validity" of these indicators. We argue that "construct validity" is not a useful tool to assess the merits of the WGI, and even if it were, Thomas provides no evidence of any practical consequences of failure to meet the criteria of construct validity.

1. Introduction

In the 1996 hit movie "Jerry Maguire", Cuba Gooding Jr. plays an up-and-coming football player who is disappointed in the inability of his agent (played by Tom Cruise) to deliver the goods in the form of a contract with a professional team. In one famous scene, Gooding calls Cruise and exhorts him to "*show me the money"*, or else he will be fired as Gooding's agent.

Thomas (2009) offers a variety of critiques of the Worldwide Governance Indicators (WGI) that we have been producing since the late 1990s.[2] Here we do not attempt to respond in detail to each and every one of the points raised by Thomas in her critique. Rather, we focus on her headline claim that the WGI do not satisfy the requirements of construct validity. In this brief response we reiterate two basic points we made in an earlier response to Thomas: we do not think that the machinery of construct validity provides useful criteria for judging the WGI, and even if construct validity did provide useful criteria, Thomas does not demonstrate any practical consequences of failure to meet these criteria.[3] Like Tom Cruise, Thomas' critique fails to "show us the money".

2. Construct Validity and the WGI

Thomas' main argument is that the WGI lack what is referred to in some areas of social science research as "construct validity", and so their use by policymakers is "arbitrary". To understand this rather sweeping (and as we argue, entirely unsubstantiated) critique, which is likely to be unfamiliar to

---

[1] Brookings Institution, World Bank Research Department, and World Bank Institute, respectively. The views expressed here do not reflect those of the Brookings Institution, the World Bank, its Executive Directors, or the countries they represent.

[2] For more details on the WGI project, including full access to the aggregate indicators and underlying source data, as well as complete documentation, see www.govindicators.org.

[3] See Kaufmann, Kraay, and Mastruzzi (2007) for a response to an earlier draft of Thomas' paper, containing further discussion of the points raised here.

many outside the few particular disciplines that use this concept, it is useful to provide some background.[4] The idea of construct validity first emerged in psychology in the 1950s as a means of disciplining the process of defining and measuring personality traits in humans. It has since spread to some other disciplines, although notably not to economics, as a tool for judging the usefulness of abstract concepts and their empirical proxies.[5]

There are two main ingredients to construct validity. The first is that the construct to be measured, in our case, concepts such as "voice and accountability", "rule of law", or "control of corruption" are clearly defined in a way that is consistent with some underlying theory. This is referred to as "content validity". In a nutshell, Thomas' critique here is that she thinks the definitions of governance that we offer are insufficiently grounded in theory. The second requirement of construct validity is that the indicators satisfy "convergent validity", i.e. the indicators are correlated with things that theory says they should be correlated with, and "discriminant validity", i.e. the indicators are *not* correlated with things that theory says they should *not* be correlated with. As we discuss below, our view is that these criteria are of dubious value in assessing the quality of empirical measures because of the confirmatory biases that they impart to subsequent empirical analysis.

3. Content Validity and the WGI

Thomas objects to our "Voice and Accountability" indicator because its definition does not coincide with that of another scholar (in this case, Hirschman (1970)), and similarly she objects to our "Rule of Law" definition because it does not coincide with another scholars' definition (in this case Fallon and Richard (1997)). While we accept that reasonable people can differ in their theoretical conceptions or definitions of governance, we think this alone is a flimsy basis for criticism of the WGI. For better or worse, it is widely acknowledged by scholars and practitioners that the term "governance", and some of its related dimensions, lack universally-accepted common definitions.[6] For example, even popular outlets such as The Economist magazine have devoted pages to discussing the variety of definitions of concepts such as the "rule of law".[7]

In the absence of such definitional consensus, we think it is entirely reasonable for us to propose our own definitions of the six governance indicators that comprise the WGI, of course basing them broadly as we did on existing definitions or understandings of the concepts. And for some of our indicators, we think the definitional ambiguities are minimal in any case. One of the most widely used WGI measures, Control of Corruption, is very clearly based on the accepted definition of the term as the use of public office for private gain. It is completely unwarranted to dismiss these indicators as "elaborate untested hypotheses" (as Thomas does in her introduction) merely due to what in our view are no more than definitional quibbles. In fact, if we were to take Thomas' critique seriously, the

[4] See Trochim (2006) for a useful survey.

[5] We find it very telling that the definitive New Palgrave Dictionary of Economics nowhere contains the term "construct validity" (see www.dictionaryofeconomics.com).

[6] In Kaufmann and Kraay (2008) we review the diversity of some of the leading definitions.

[7] See The Economist. "Order in the Jungle". March 13, 2008.

absence of definitional consensus in the profession regarding governance would paralyze any effort to measure governance using any means.

We accept fully that other scholars, including of course Thomas, may well disagree with the definitions of governance that we have offered, and they are perfectly entitled to do so. However, as we noted elsewhere (Kaufmann, Kraay and Mastruzzi (2007)), there is "free entry" in the market for governance indicators. If Thomas has different or better theoretical and/or definitional conceptions of governance than we do, then she can articulate them, construct corresponding alternative empirical measures, show that they differ in important ways from what we have thus far done, and by doing so convince the profession that her conceptualization should be used instead of, or as a complement to, the WGI. Instead, all Thomas does in her critique is offer a few slightly different definitions of two of the dimensions of governance that we measure, and provides no evidence of how these alternative definitions would result in significantly different empirical measures. This first part of Thomas' critique based on content validity, like Tom Cruise, fails to "show us the money".

4. Convergent and Discriminant Validity and the WGI

The second requirement of construct validity is that the indicators satisfy convergent validity and discriminant validity. Our view is that these criteria are of dubious value in assessing the quality of empirical measures such as the WGI. To see why, imagine that our starting point were a theory positing that corruption is less prevalent in countries that rank highly in alphabetical order. Such a theory might not be so implausible. After all, it would be supported by anecdotal evidence – casual observation suggests that corruption is much worse in Zimbabwe and Zambia than it is in Australia or Austria. And conceptually one could theorize about mechanisms through which this might be true – for example, countries ranking high in alphabetical order might come up for scrutiny sooner in international organizations where countries are routinely ordered alphabetically.

With this theoretical starting point, suppose that we use the criterion of convergent validity to assess the WGI Control of Corruption (WGI-CC) measure. The correlation of WGI-CC with the alphabetical rank order of country names in 2007 is -0.06 and is not significantly different from zero, while the theory described above predicts a strong positive correlation. Thus we would reject WGI-CC as an empirical measure of the construct of corruption, on the grounds that it fails to meet convergent validity – it is not correlated with what the theory says it should be correlated with. What measure of corruption would satisfy convergent validity? We might for example propose using the alphabetical rank order of a country's three-letter World Bank country code as an alternative measure of corruption. The correlation of this alternative measure of corruption with the alphabetical rank of country names is 0.85, and so the criterion of convergent validity would suggest we prefer this alternative measure of corruption. [8]

This example is obviously frivolous. But it illustrates what we think is a fundamental flaw in using the criteria of convergent and discriminant validity to choose indicators: that of confirmatory bias,

[8] This correlation is not perfect because World Bank country codes do not always start with the same letter as the name of the country in English, for example the country code for Chad is TCD.

in the sense that convergent and discriminant validity force users to select empirical measures that are consistent with prior theorizing. But in our view this undermines the primary purpose of data – which is to test theories to assess their empirical relevance. To give a less frivolous example, suppose one reasonably theorized that corruption mattered for development. Strict application of convergent validity would lead us to prefer those measures of corruption that are more strongly correlated with per capita income across countries. Tests of the theory that corruption matters for development would automatically deliver the desired result. We think this is why the machinery of construct validity on which Thomas bases her critique of the WGI is not used in our profession which takes seriously the use of data to distinguish between competing hypotheses.

Even if one were to take the criteria of convergent and discriminant validity seriously, Thomas' critique fails to demonstrate any evidence of success or failure of the WGI with respect to these criteria. The six aggregate indicators of governance that we have constructed, together with virtually all of the underlying source data, are freely available through our website (www.govindicators.org). It would be straightforward for Thomas to identify a set of variables that she thinks should be correlated with our governance indicators according to her preferred theoretical worldview, and another set of variables that should not be, and then use them to empirically document the success or failure of the WGI with respect to the criteria of convergent and discriminant validity. We find it very telling that she has not done so during the three years that her critique has been under development. On this point as well, Thomas' critique fails to "show us the money".

References

Hirschman, A. O. (1970) "Exit, Voice, and Loyalty". Cambridge: Harvard University Press.

Fallon, J. and H. Richards (1997). "The Rule of Law as a Concept in Constitutional Discourse". Columbia Law Review. 97(1):1-56.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2007). "Answering the Critics". World Bank Policy Research Department Working Paper No. 4149.

Kaufmann, Daniel and Aart Kraay (2008). "Governance Indicators: Where Are We, and Where Should We Be Going?". World Bank Research Observer. 23:1-30 (Spring 2008).

Trochim, William (2006). "Construct Validity" (http://www.socialresearchmethods.net/kb/constval.htm)