

1

Transparency in Measurement

This book is the first of four books theorizing the structure of governance above and below the central state. We describe the theory as postfunctionalist because it claims that governance, which we define broadly as authoritative decision making in the public sphere, is determined not just by its functionality but by its emotional resonance. Multilevel governance within the state, the topic of this book, evokes intense preferences not just for what it *does*, but for what it *is*. Jurisdictional design has intrinsic meaning for people. It expresses their national, regional, and local identities. The premise of post-functionalism is that this cannot be reduced to the extrinsic functions of governance. It is about “who are we” as well as “who gets what.”

This raises questions that can be answered only by looking within countries. Over the past two decades there has been an upsurge of research on territorial governance within countries, but measurement has lagged behind. Case studies investigate the mobilization of ethnic minorities and the efforts of central rulers to accommodate or suppress them, but the effects are only dimly perceived in national indices, and they escape fiscal measures entirely.

A measure is a disciplined summary. It attaches conceptual relevance to some phenomena and ignores others. As one begins to conceptualize variation in territorial governance, one enters a subterranean world in which there are numberless possibilities. Jurisdictional regions vary enormously in size and population. Their authority varies more than that of states. Some are merely central outposts for conveying and retrieving information. Others exert more influence over the lives of people living under their rule than the national state itself. One must leave behind the idea that territorial governance is constitutionalized, and therefore highly stable. Regional governance is governance in motion. The regional authority index detects more than 1300 changes in sixty-two countries.¹

¹ This is the number of changes of 0.1 or more on one of the ten dimensions for a region or regional tier.

Transparency in Measurement

Thirty-four new tiers of regional governance have been set up and seven have been abolished. Precise observation of territorial governance reveals a landscape that is fascinating in its flux and diversity.

This book sets out a measure of regional authority that can be used by social scientists to investigate the character, causes, and consequences of governance within the state. In this chapter we explain the key decisions that underpin our measure. How do we conceptualize regional authority? How do we summarize this abstract concept in dimensions? What indicators do we use to tap variation along these dimensions? And how do we score cases using these indicators? Each step is a theoretically motivated move from the abstract to the concrete. Subsequent chapters allow the reader to assess the validity of these steps and of the final product. Chapter Two compares our measure with other commonly used measures of decentralization. Chapter Three is a hands-on guide to the rules underpinning the measure and its indicators. The book concludes with profiles that overview change in regional authority across eighty-one countries on a common analytical frame.

We have three purposes. First, we wish to provide a reasonably valid measure of subnational government structure that is sensitive to cross-sectional and temporal variation. The measure conceives subnational governance as a multidimensional phenomenon that can take place at multiple scales. Fiscal measures provide annual data for a wide range of countries, but the amount of money that passes through a subnational government may not accurately reflect its authority to tax or spend. And there is much more to the structure of government than spending or taxation. Some regional governments can block constitutional change; some control local government, immigration, or the police; some play an important role in co-governing the country as a whole. The concept of federalism does a better job at capturing regional authority, but it is insensitive to reform short of constitutional change and does not pick up cross-sectional variation among federal or among unitary countries. The measure proposed here detects a lot of variation both within these categories and over time. The figures preceding the country profiles reveal that the territorial structure of government is much more malleable than is implied by the classics of comparative politics (e.g. Lijphart 1999; Riker 1964).

Our second purpose is to break open subnational government so that others may look inside. Comparative politics is conventionally seen as the study of politics across countries. Still, the field has a prominent and longstanding tradition of studying politics not just across, but also within, countries. Among the most celebrated examples are Tocqueville's *Democracy in America* (1838), which compares American states to assess the effects of slavery, Seymour Martin Lipset's *Agrarian Socialism* (1950), which compares wheat-belt provinces in Canada and the US, and O'Donnell's (1973) discussion of regions in Argentina and Brazil.

Measurement

The past decade has witnessed an upsurge in the number of articles and books comparing regions within and across countries.² The most obvious reason is that we live in an era in which authority has spun away from central states to subnational and supranational governments. We see this very clearly in our measure, which reveals an increase in the authority of regional governments in two-thirds (fifty-two) of the countries we observe. Another reason is that comparing regions can lead to better causal inference. Democracy, economic growth, crime, and many other things that people care about, vary within as well as among countries (Snyder 2001; Giraudy 2015; Giraudy, Moncada, and Snyder 2014).³

Subnational comparison can increase the number of relevant observations. More importantly, it can provide inferential leverage in engaging the fundamental problem with observational data: too much varies and the controls one can impose through matching and fixed effects are both demanding and incomplete. This is where subnational comparison is particularly useful. Many of the confounding factors that are difficult to control for are national, and controlling for national factors is a powerful lever for explaining variation against a background of commonality. This is precisely Robert Putnam's inferential strategy in *Making Democracy Work* (1993). Comparing regions in the north and south of Italy allows him to control for a wide array of factors—including Catholicism, parliamentarism, and the legacy of fascism—that could plausibly influence democratic performance.

This calls for measurement at the level of the individual region rather than the country—a decision that has shaped every aspect of this book. Examining territorial government inside countries brings to life phenomena that are otherwise invisible. More than half of the countries with a population greater than twenty million have not one, but two or more levels of intermediate government. An increasing number of countries are differentiated, that is, they have one or more regions that stand out from other regions. We wish to compare not just countries, but regions and regional tiers within countries. And we compare not only how regional governments exert authority over those living in its territory, but also how they co-govern the country as a whole. In short, the question we are asking is “In what ways, and to what extent, does a regional government possess authority over whom at what time?”

² This trend encompasses Western countries (e.g. Dandoy and Schakel, eds. 2013; Gerring, Plamer, Teorell, and Zarecki 2015; Kelemen and Teo 2014; Kleider 2014), Latin America (e.g. Giraudy 2015; Chapman Osterkatz 2013; Niedzwiecki 2014), Africa (Posner 2004), Russia (Robertson 2011), and China (Landry 2008; Tsai 2007).

³ For studies that are explicitly motivated by this insight, see e.g. Agnew (2014); Charron and Lapuente (2012); Gibson (2012); Harbers and Ingram (2014).

Transparency in Measurement

The product is a measure that provides information on the financial, legal, policy, representational, and constitutional competences of individual regions and regional tiers on an annual basis. Each of the ten dimensions of the measure picks up a distinct component of regional authority.⁴ We aggregate dimensional scores for regions and tiers to the country level, but researchers can re-assemble the constituent dimensions for their own purposes. They can also begin to examine the effects of variation in the way in which regional governments exert authority. Why, for example, do some regional governments exercise considerable powers within their own borders, but have almost no role in governing the country? What is the effect of tying a region into country-wide governance? How do fiscal, legal, policy, representational, and constitutional competences interact, and with what results? Why has subnational governance become more differentiated over time?

Our third purpose relates to measurement in general. How should one go about measuring a big abstract concept such as authority? In our 2010 book which introduced the regional authority index (RAI) we emphasized that it was vital to lay our method bare before the reader “so that others may replicate, amend, or refute our decisions” (Hooghe, Marks, and Schakel 2010: 3). We wanted to make it possible for others to evaluate how the measure was constructed, and we were intensely aware that our decisions were theory-driven. This is the commitment to transparency that has been set out by the American Political Science Association in a series of collectively authored statements. Beyond the well-recognized (though not always practiced) norm that researchers provide access to the data and analytical methods they use in their publications, the APSA (2012: 10) calls for production transparency: “Researchers providing access to data they themselves generated or collected, should offer a full account of the procedures used to collect or generate the data.”

Production transparency implies providing information about how the data were generated or collected, including a record of decisions the scholar made in the course of transforming their labor and capital into data points and similar recorded observations. In order for data to be understandable and effectively interpretable by other scholars, whether for replication or secondary analysis, they should be accompanied by comprehensive documentation and metadata detailing the context of data collection, and the processes employed to generate/collect the data.

⁴ The financial statistics produced by the International Monetary Fund (IMF) are a model worth emulating. The data take the form of a multidimensional matrix which breaks down financial flows by type of transaction, institutional unit, sector, and as discussed later, by jurisdictional level. “In contrast to summary measures, the detailed data of the GFS [Government Finance Statistics] system can be used to examine specific areas of government operation. For example, one might want information about particular forms of taxation, the level of expense incurred on a type of social service, or the amount of government borrowing from the banking system” (IMF 2014: 3). The RAI consists of ten dimensions and a larger number of indicators that can be individually analyzed and re-aggregated.

Measurement

Production transparency should be thought of as a prerequisite for the content of one scholar's data to be truly accessible to other researchers. Analytic transparency is a separate but closely associated concept. Scholars making evidence-based knowledge claims should provide a full account of how they drew their conclusions, clearly mapping the path on the data to the claims (Lupia and Alter 2014: 57, citing a memo by Lupia and Elman 2010).

Production transparency is a public good that lies at the heart of the scientific method. Science operates by the light of day, by making the process of confirmation and disconfirmation explicit. This applies as much to measurement as to the methods used to analyze data. Estimating a political concept requires a series of theoretical, conceptual, operational, and coding decisions. Each step is a move from the general to the particular in which an abstract concept is translated into the language of numbers. Measurement, no less than theory, is "the art of discerning what we may with advantage omit" (Popper 1982: 44).

The process can be broken down into six steps.

- 1) *Defining the background concept.* How have social scientists understood the concept?
- 2) *Specifying the measurement concept.* Which of those meanings does one wish to include?
- 3) *Unfolding the concept into dimensions.* How does one break down the measurement concept into discrete pieces that can be independently assessed and aggregated to capture its meaning?
- 4) *Operationalizing the dimensions.* How does one conceptualize and specify intervals on the dimensions? What rules allow one to reliably detect variation across intervals?
- 5) *Scoring cases.* What information does one use to score cases? Where is that information, and how can others gain access to it?
- 6) *Adjudicating scores.* How does one interpret gray cases, i.e. cases for which scoring involves interpretation of a rule?

Figure 1.1 is an expanded version of Adcock and Collier's (2001) schema.⁵ The arrows are verbs to describe the steps down from the background concept

⁵ We make two additions. The first is a level of measurement, *dimensions*, in which the abstract concept is broken down into components prior to developing indicators. Virtually all concepts of major theoretical interest in the social sciences are complex in that they are comprised of more than a single dimension of variation. So an important step in operationalizing abstract concepts such as regional authority, democracy, or gross national product (GNP) is to conceive a limited set of dimensions that are amenable to operationalization and that together summarize the meaning of the overarching concept. The second addition is a final important step, *adjudicating scores*, which lays out rules for exceptional or difficult cases that arise in any coding scheme. Social science measurement is replete with gray cases, and one telling indication of the transparency of a measure is whether these are explicitly communicated.

Transparency in Measurement

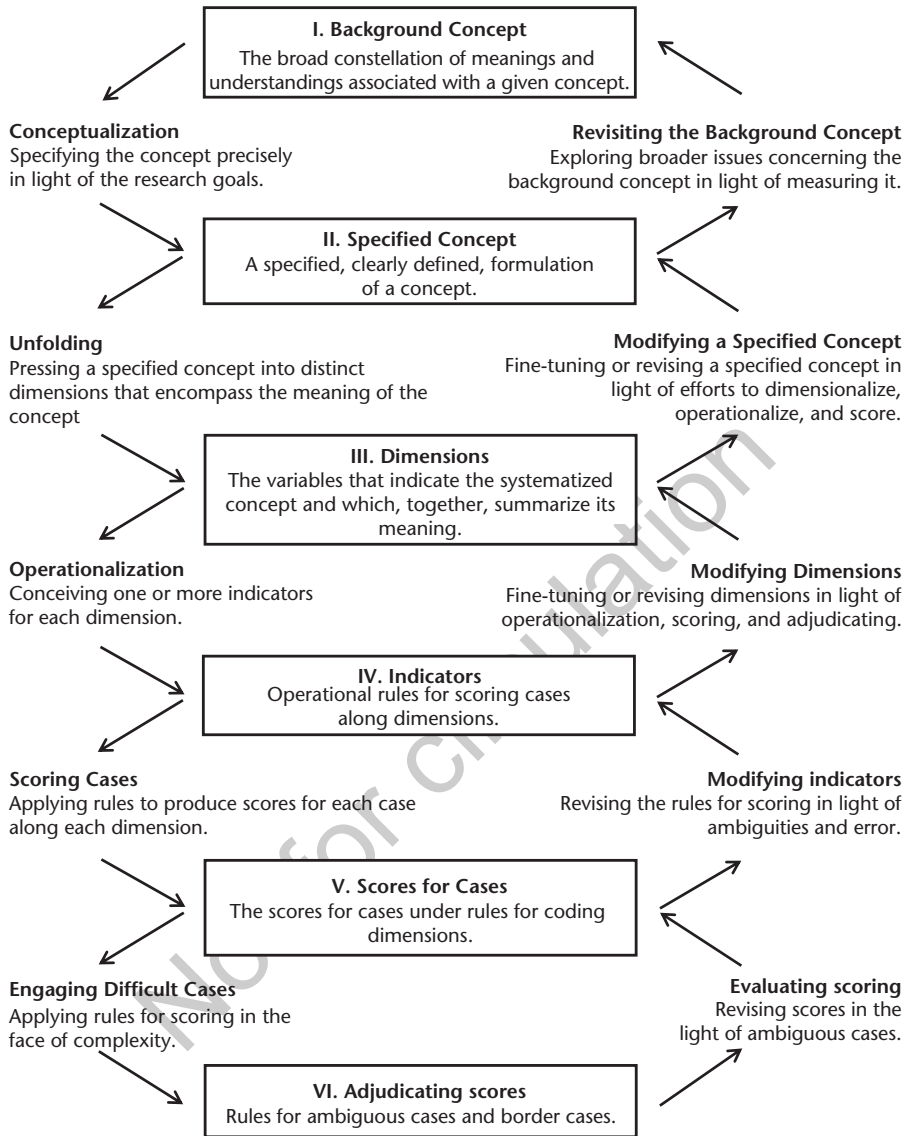


Figure 1.1. Measurement model

to individual scores or up from individual scores to the background concept. The boxes contain nouns to describe the concept, its dimensions, indicators, and scores as one presses the concept closer to phenomena that can be observed at lower levels of abstraction. The figure makes the point that these steps are interdependent. How one specifies the scope of a concept has consequences for breaking it into dimensions. How one operationalizes those

dimensions frames the choice of appropriate indicators. Even minor differences in the indicators can have serious consequences for scoring.

Making this transparent is good for several reasons. Transparency facilitates replication. It is true that we rarely replicate each other's results, but the possibility of replication has an effect on the quality of science that reaches beyond its incidence. Most findings will never be replicated, but the more influential a finding, the greater the likelihood it will be replicated. Replication is insurance for Richard Feynman's (1985: 343) first principle of science: "[Y]ou must not fool yourself, and you are the easiest person to fool."⁶

Transparency allows others to understand and probe the inner workings of a measure, and this can help in assessing its validity. A dataset is a matrix of decisions that cannot, even in principle, be inducted from the numbers that appear in the cells. One must have access to those decisions to assess the numbers. Transparency directs attention to the construction of a measure, and exposes the decisions that underpin it. I have little direct knowledge of how the gross domestic product (GDP) of the US grew in the last quarter, but I do have direct knowledge of the process by which the data were collected (Landefeld et al. 2008). I have little direct knowledge of the people who, in the week of September 8, 2015, intended to vote in favor of Scottish independence, but I do know (or should know) how a survey instrument was constructed, how the population was sampled, and how the survey was conducted.

However, transparency can do more than tell one how a measure is produced. It can allow others to evaluate the validity of the scores for individual cases. We can be reasonably sure that some experts will know more about the structure of government in their country than we will ever know. Transparency can reveal the evidence and reasoning that go into individual scores. Let others see how one arrives at particular scores for cases with which they are deeply knowledgeable. Let them have access to the judgments that produce scores for gray cases. This is why we devote considerable space to country-specific profiles that provide an overview of regional governance and explain how we score particular regional reforms in a country. Explaining the construction of a measure and investigating its reliability are not at all the same as explaining how individual cases are scored. However, it is the scores for individual cases that are of most use-value. The profiles provide a birds-eye view of regional governance across a wide range of countries on a common

⁶ Or, as Alexander Pope ([1734] 1903: 157) wrote

To observations which ourselves we make,
We grow more partial for th' observer's sake.

Transparency in Measurement

format. By making our judgments explicit we can ask experts: “Have we used the appropriate evidence?” “Do our judgments make sense?”

Measurement, Error, and Fallibility

Measurement is inherently prone to error. This is the thrust of Lakatos’ philosophy of scientific method, which rejects the demarcation of measurement and theory (Lakatos 1970; see also Bouwmans 2005). Measurement maps a property of the empirical world onto a set of numbers, a procedure that requires a series of inferential steps. In the words of a contemporary philosopher of science: “Measurement involves a host of theoretical and statistical representations of measuring systems and the data they produce” (Tal 2013: 1164). Social scientific measurement is at least as inferentially complex as measurement in the physical sciences, so it is worth taking epistemologists seriously when they point out that “physicists are forced to test the theories of physics on the basis of the theories of physics” (Chang 2004: 221). An observation is a theoretically guided experiment that produces information by making claims about what is observed and how it is observed. The philosopher–scientist Pierre Duhem ([1906] 1954: 182) stresses that “it is impossible to leave outside the laboratory door the theory we wish to test, for without theory it is impossible to regulate a single instrument or to interpret a single reading.”

The appearance of hard facts is deceptive even in the measurement of something as basic as temperature. Comparing temperature observations in different places called for some well defined fixed points. The temperature of the human body and that of the cellar in the Paris observatory provided useful (but not entirely reliable) fixed points until Anders Celsius created a universal scale using the boiling point and freezing point of water. Evidently Celsius conceived his scale as a measure of degrees of cold, not heat. Water boiled at 0° on Celsius’ original scale, and froze at 100° (Beckman 1997; Chang 2004: 159ff). Early thermometers used either alcohol or mercury. But the premise that alcohol and mercury thermometers could be made to “speak the same language” was disconfirmed when Réaumur found that recalibration from one to the other failed to produce uniform readings (Gausson 1739: 133; Réaumur 1739; Chang 2001). Mercury became the standard because the rate at which it expanded approximated the ratios of mixing ice and boiling water.⁷ However, this assumed that mercury thermometers would give uniform readings if they were made of different kinds of glass, and more fundamentally, it assumed

⁷ Or, more precisely, nearly freezing and nearly boiling water (Chang 2004: note 27).

Measurement

that the temperature is an additive function of the ratio of freezing and boiling water. When the linear theory of mixing was disconfirmed, thermometers used gas on the ground that the molecular interactions that produced non-linearity in liquids would be nearly absent in gas. Gas thermometers were accurate for most purposes, although the technology has moved on and the current International Temperature Scale has the boiling point of water at atmospheric pressure as 99.975°C rather than 100°C.

Having an accurate thermometer is just the first step in reliably measuring global temperature.⁸ Many measurement stations are located near population centers that are warmer than the surrounding areas. Irrigation has the opposite effect. The coverage of many parts of the globe, including particularly the hottest and coldest regions, is incomplete. Not only are estimates inexact, but there are numerous sources of systematic bias. Ships now measure ocean surface temperature with water flowing through engine cooling water intakes rather than with water collected in buckets (Matthews 2013). The introduction of the new method coincides with a rise in ocean temperature in the 1940s, perhaps because water collected in buckets cooled prior to measurement. Social factors come into play. Daily mean temperatures are calculated by summing the maximum and minimum over a twenty-four-hour period and dividing by two. However, volunteer weather observers have an understandable reluctance to take midnight readings, and until the 1940s most weather stations recorded the maximum and minimum temperatures for the twenty-four hours ending near sunset (Karl et al. 1986). Scientists seek to correct these and other possible sources of bias using proxies such as satellite measurement of the intensity of night light to adjust for the urban heating effect. None of these potential biases is large enough to shake the inference that global warming is taking place, but they do lead an expert inquiry to emphasize that on account of urbanization and observational irregularity, “Temperature records in the United States are especially prone to uncertainty” (Hansen et al. 2010: 103).

No less than in the physical sciences, measurements in the social sciences are based on a series of inferences, each of which can be questioned. The general lesson is that no observation can sit in judgment of a theory without being cross-examined. And there is no reason why the interrogation of an observation should be less searching than the interrogation of a theory. The implication that Lakatos draws from this is that “clashes between theories and factual propositions are not ‘falsifications’ but merely inconsistencies. Our imagination may play a greater role in the formulation of ‘theories’ than in the formulation of ‘factual propositions’, but both are fallible” (Lakatos 1970: 99–100).

⁸ We thank Michaël Tatham for drawing our attention to this.

Transparency in Measurement

All observation is fallible, but some observations are more fallible than others. Social scientists are concerned with highly abstract concepts, many of which have normative connotations. The chain of inference that links the observation of a particular behavior to the concept of democracy, inequality, or decentralization is both long and complex. In this endeavor the assumption that measurement error is random rather than systematic is false comfort, for it suggests that issues of validity can be reduced to issues of reliability. One of the purposes of observation is precisely to discipline our theories or “guesses” (Feynman 1965: 156). However, this takes the form of a conversation rather than a judgment, for the observations that one brings to bear are themselves built on a scaffold of theoretically motivated short-cuts.

Perhaps in no other field of political or economic science is this more apparent than in the study of the structure of government, and decentralization in particular. Theoretical expectations often line up on both sides of the street, but the information that is used to test them can be slippery. Weak theory and poor measurement are complementary because almost any set of observations appears consistent with one or another theory. Summarizing the effects of decentralization for economic performance and the quality of government, Treisman (2007: 5) writes that “as one would expect given the uncertain and conditional results of theory, almost no robust empirical findings have been reported about the consequences of decentralization.”

An extensive literature takes up the question of the effect of decentralization on the size of the public sector. This is the “Leviathan” question introduced by Brennan and Buchanan (1980): Is government intrusion in the economy smaller when the public sector is decentralized? Brennan and Buchanan argue that it is, but others have developed plausible models that claim exactly the opposite (e.g. Oates 1985; Stein 1999). Intervening variables can change the sign of the effect. Oates (2005) argues that “it is not fiscal decentralization *per se* that matters, but what form it takes” (Oates 2005; Rodden 2003a; Jin and Zou 2002).

The standard measure of decentralization in this literature is World Bank data derived from the IMF’s Government Finance Statistics (GFS) measuring subnational expenditures or subnational revenues as a proportion of total government expenditures or revenues.⁹ Data are rarely reported for the two tiers of subnational government in the GFS framework, and the criteria for intermediate and local government vary across countries. Several countries, including France, Italy, Japan, the Philippines, and Thailand, have no

⁹ The World Bank is explicit about the limitations of these data: “Shared taxes appear as sub-national revenue, although the sub-national government has no autonomy in determining the revenue base or rate, since the GFS reports revenues based on which level of government ultimately receives the revenues.” <<http://www1.worldbank.org/publicsector/decentralization/fiscalindicators.htm#Strengths>>.

Measurement

intermediate tier of government in the dataset because their regions are reported as local government. Belgian communities, which form one the strongest intermediate levels of government anywhere, are classed as part of central government with the result that Belgium comes out as the most centralized country in the Organisation for Economic Co-operation and Development (OECD).¹⁰

But the more fundamental issue is conceptual. Is the amount of money that a subnational government raises or spends a valid measure of decentralization (Rodden 2003a)? If decentralization involves the authority to make decisions, the answer must be “not necessarily.” Sweden, Norway, Finland, and Denmark—big spending governments with a history of social democratic rule—are considered to be highly decentralized because they channel considerable funds through their local governments. However, local governments in these countries spend and tax according to national laws (see Chapter Two). The IMF data consider these countries, on average, to be as decentralized as the US or Germany, and more than twice as decentralized as Spain, Italy, or France. Perhaps not surprisingly, a recent paper using these data concludes that “fiscal decentralization leads to larger public sectors when the federal government is controlled by a left-wing party, and to smaller public sectors when it is controlled by a right-wing party” (Baskaran 2011: 500).

The most commonly used alternative measure in the Leviathan literature is a dichotomous variable that distinguishes federal from non-federal countries. This has surface validity, but it is useful only in cross-sectional analysis because few countries cross the federal divide. This variable also censors variation within each category. Non-federal countries include both highly centralized countries, such as El Salvador and Luxembourg, and countries, such as Indonesia and Spain, which in our data are more decentralized than several federal countries. Knowing whether a study uses this federalism variable or IMF fiscal data helps one predict whether that study confirms or disconfirms the hypothesis that decentralization reduces public spending. A meta-analysis (Yeung 2009: 22) concludes that “Despite over 36 years of research, little consensus has emerged on the effect of fiscal decentralization on the size of government” and that the reasons for disagreement have to do with theoretical and conceptual choices that are implicit in “a study’s unit of analysis and measure of decentralization.”

Every measure produces information by making theoretical and conceptual claims about the world. A measure of regional authority can no more be insulated from theory than a measure of temperature. Neither theory nor data can sit in judgment on the other. Rather they need to be brought into a

¹⁰ Similarly, Scotland and Wales are assessed as part of the UK central government (IMF 2008: 546).

Transparency in Measurement

dialogue in which each is regarded as fallible. Charles Darwin, who spent much of his life making careful observations, remarked that “a good observer really means a good theorist” (Darwin 1903: 82).

Nuts and Bolts

We seek to measure the authority exercised by regional governments in eighty-one countries on an annual basis from 1950, or from the time a country becomes independent, to 2010.¹¹ The sample consists of all European Union (EU) member states, all member states of the OECD, all Latin American countries, ten countries in Europe beyond the EU, and eleven in the Pacific and South-East Asia.¹²

Table 1.1 lists four prior measures of regional authority by year of publication. Measurement has become more comprehensive over time, providing more information for more years. The measure set out here continues this development and has some unique features.

Most importantly, the unit of analysis is the individual region, which we define as a jurisdiction between national government and local government.

Table 1.1. Measures of regional authority

	Lijphart (1999)	Woldendorp, Keman & Budge (2000)	Arzaghi & Henderson (2005)	Brancati (2008)	Regional Authority Index (2016)
Country coverage	36 pre-1990 Western democracies	37 Balkan, OECD, EU democracies	48 countries with population > 10 million	37 countries with regional ethnic groups	81 Western, post- communist, Latin American, Southeast Asian & Pacific countries
Time coverage	1945–1996	1945–1988	1960–1995	1985–2000	1950–2010
Time points	1	1	8	16	61
Individual regions	no	no	no	no	yes
Multiple tiers	no	no	no	no	yes
Observations per country/year	5	4	8	5	10–130

¹¹ On average a country in the dataset is coded for forty-seven years. Forty-eight countries are coded for the entire 1950–2010 period.

¹² The case selection reflects a trade-off between an effort to cover the largest possible number of countries and the team’s resources—chiefly their time—and the availability of sources and country expertise.

Measurement

We draw the boundary between local and regional government at an average population level of 150,000. This excludes the lowest tier of government in all eighty-one countries, but allows us to capture intermediate governments, often arrayed at two nested levels between the local and national. We relax the population criteria for individual jurisdictions, such as Greenland or the Galapagos islands, that stick out from a tier of government that meets the regional threshold.

A focus on regional or intermediate government has some theoretical and practical virtues. It encompasses virtually all subnational governments that exert self-rule within distinct homelands. Such governments tend to form part of a regional tier of government with an average population greater than 150,000 or they have special authoritative competences alongside a regional tier. Where subnational governments play an important role in co-governing a country, these are almost always intermediate governments. To the extent that subnational governments play a formally recognized role in shaping constitutional reform, one needs, again, to look to the intermediate level. Yet many countries lack any form of intermediate governance or have regional governments that are merely deconcentrated. Regional jurisdictions are the most variable elements of territorial governance within the state and are generally the most contested.

The decision to conceptualize the individual region as the unit of analysis has several consequences. It raises the possibility that regions may be nested within each other at different scales. Altogether, there are 103 levels of regional government in the sixty-five countries that have at least one tier of regional government. So researchers can begin to compare regional tiers within countries. The measure picks up reform even when it is limited to a single region in a country. A reform in a single region may not seem much, but if it undermines the norm that all be treated equally, it may be hotly contested by other regions as well as the central government. Moreover, such a reform may threaten the break-up of the state.

The measure comprises ten dimensions that tap the diverse ways in which a region may exert authority. These dimensions are quite strongly associated with each other and can be thought of as indicators of a latent variable. Yet those who are interested in examining the pathways to regional authority can disaggregate regional authority into its components. Some dimensions, including those that tap regional representation, policy scope, and borrowing autonomy, exhibit more reform than others.

Combining a regional approach with fine grained attention to the ways in which a region can exert authority produces a measure that is considerably more sensitive to change than any previous one. Twenty-one percent of the variation occurs over time. The territorial structure of governance is much less fixed than one would assume when reading the classics of comparative

Transparency in Measurement

politics such as Arend Lijphart's *Patterns of Government* (1999) or Daniel Elazar's *Exploring Federalism* (1987).

However, the RAI is limited in some important respects. Three stand out. We do not encompass tiers of subnational government containing jurisdictions with an average population less than 150,000. Hence, we omit local government entirely. This is a topic that calls for systematic measurement, perhaps adapting the measure proposed here to variation in the policy responsibilities of local authorities (Campbell 2003; Loughlin, Hendriks, and Lidström 2011; Nickson 2011; Norton 1994; Page and Goldsmith 1987, 2010).¹³

The regional authority index excludes informal arrangements. It is concerned exclusively with authority, which we define as formal power expressed in legal rules. Hence it omits contextual factors, such as leadership, political parties, or corruption, which may affect government performance. Finally, the country coverage of the present measure is incomplete. In particular, it does not cover China or India, two continental sized countries with correspondingly complex and differentiated systems of regional government.

I. *The Background Concept: Political Authority*

Political authority is a core concern of political science, some would argue *the* core concern (Eckstein 1973; Lake 2010; Parsons 1963; Weber 1968). Political authority—the capacity to make legitimate and binding decisions for a collectivity—underpins human cooperation among large groups of individuals. Human beings cooperate in order to produce goods that they could not produce individually. These goods include law, knowledge, and security. These goods are social in that they benefit all who live in the collectivity, and they are inclusive in that their benefits cannot practically be limited to those who contribute for them.¹⁴ Whereas small communities can impose social sanctions to produce public goods, large groups are far more vulnerable. The exercise of political authority diminishes the temptation to defect from collective decisions, and reassures those who do cooperate that they are not being exploited: “For although men [in a well ordered society] know that they share a common sense of justice and that each wants to adhere to the existing arrangements, they may nevertheless lack full confidence in one another. They suspect that some are not doing their part, and so they may be tempted not to do theirs” (Rawls 1971: 211).

Authority is relational: *A* has authority over *B* with respect to some set of actions, *C*. This parallels Robert Dahl's (1957: 202–3, 1968) conceptualization

¹³ A team led by Andreas Ladner and Nicholas Keuffe is adapting the RAI to estimate local decentralization in thirty-eight countries (personal communication, March 2015).

¹⁴ The negative formulation is that public goods are non-excludable and non-rivalrous.

of power as the ability of *A* to get *B* to do something that *B* would not otherwise do. A short-hand definition of authority is *legitimate* power. One speaks of authority if *B* regards *A*'s command as legitimate and correspondingly has an obligation to obey. Authority implies power, but power does not imply authority. Whereas power is evidenced in its effects irrespective of their cause, authority exists only to the extent that *B* recognizes an obligation resting on the legitimacy of *A*'s command. Such recognition may have diverse sources, including charisma, tradition, and religion (Weber 1958). This book is concerned with the modern variant of authority—legal-rational domination based in a codified legal order.

Two conceptions have predominated in our understanding of the structure of authority. The first conceives a polity as grounded in human sociality. Families, villages, towns, provinces, and other small or medium scale communities are the ingredients of larger political formations. This idea is as close to a universal principle in the study of politics as one is likely to find. Ancient states and tribes were composed of demes, wards, or villages. Aristotle conceived the polis as a double composite: households within villages; villages within the polis. Each had a collective purpose and a sphere of autonomy. The Romans built a composite empire by attaching a vanquished tribe or polis by a *foedus*—a treaty providing self-rule and protection and demanding payment of a tax, usually in the form of manpower (Marks 2012). The Qin dynasty that united China in 221BC had a four-tiered structure extending from the family through wards and provinces to the empire (Chang 2007: 64). The Incas conceived of five hierarchically nested tiers reaching from the family to an empire encompassing much of contemporary Peru, Bolivia, Ecuador, and northern Chile (Rowe 1982). Medieval scholars conceived the state as a composite (*consociandi*) of men already combined in social groups (*sympiotēs*). Johannes Althusius (1997 [1603]) conceived the state as a contract among such associations, a *consociatio consociationum* consisting of families within collegia within local communities within provinces.

The modern variant of this idea is federalism, which describes a polity “compounded of equal confederates who come together freely and retain their respective integrities even as they are bound in a common whole” (Elazar 1987: 4). Federalism highlights the basic constitutional choice between a unitary and federal system. A unitary system has a central sovereign that exercises authority, whereas a federal system disperses authority between “regional governments and a central government in such a way that each kind of government has some activities on which it makes final decisions” (Riker 1987: 101; Dahl 1986: 114). Most importantly, regions or their representatives can veto constitutional reform. The unitary/federal distinction informs a literature on the political consequences of basic constitutional decisions, including particularly ethnic conflict (Amoretti and Bermeo 2004;

Transparency in Measurement

Lijphart 1999). Federalist scholars have told us a lot about why independent units would wish to merge and how some polities arrive at federalism in order to avoid falling apart (Rector 2009; Roeder 2007; Stepan, Linz, and Yadav 2011). And there is a rich literature comparing federal polities (Watts 1998, 1999a, 2008).¹⁵

The federal/unitary distinction draws attention to the tension between self-rule and shared rule that is inherent in a composite polity. The constituent communities wish to retain their independence, their distinct way of life, their language, religion, dress, customs, their norms of social interaction. Yet they wish also to gain the benefits of scale in security, trade, and governance by forming a state in which they share rule with the center. As we discuss later, the concepts of self-rule and shared rule motivate our measurement scheme, and they are taken directly from the federalism literature.

However, the unitary/federal distinction has some fundamental limitations for the measure we propose. It is a blunt instrument for assessing incremental institutional change. Shifting from a unitary to a federal regime (or the reverse) is a high hurdle that few countries meet. The number of federal countries in our dataset has hardly changed over the past sixty years, yet there is ample evidence that this has been a period of profound reform.¹⁶ Not surprisingly, the federalism literature tells one far less about variation among unitary countries than among federal countries (Hooghe and Marks 2013; Rodden 2004; Schakel 2008). Variation among unitary countries has grown a lot over the past six decades, whereas the contrast between unitary and federal countries has diminished. Finally, federalism is concerned with the topmost level of subnational governance, whereas several countries have two or three levels of government between the national and the local.

A second conception, the idea that governance can be more or less decentralized, has also been hugely influential. Centralization and decentralization are poles of a continuous variable describing the extent to which authority is handled by the central government *versus* any government below. This way of conceiving governance is elegant and thin. Both its virtues and vices arise from its very high level of abstraction. It travels well. It allows one to compare governance around the world and over time on a single scale.

¹⁵ There has been a veritable revival in the study of federalism. Recent examples include Anderson (2012); Bednar (2009); Benz and Broschek (2013); Bolleyer (2009); Burgess (2012); Chhibber and Kollman (2004); Erk (2008); Falleti (2010); Rodden (2006); Swenden (2006); Rodden and Wibbels (2010). This wave also comprises several handbooks, such as Loughlin, Kincaid, and Swenden (2013) on federalism and regionalism, and Haider-Markel (2014) on state and local relations in the US.

¹⁶ As Gary Goertz (2006: 34) observes, dichotomous concepts tend “to downplay, if not ignore, the problems—theoretical and empirical—of the gray zone. Often, to dichotomize is to introduce measurement error . . . [because it] implies that all countries with value 1 are basically equivalent.”

We seek to develop a measure that is similarly robust across time and place. If the RAI is aggregated to the country level it can be interpreted as a measure of decentralization. We follow decentralization scholars by distinguishing forms of decentralization: over policy making; over fiscal policy; over the appointment of subnational decision makers; and over the constitution. Each can be considered an independent variable that can register change in the absence of sweeping constitutional reform.

However abstractness has a price if it comes “at the expense of connotation” (Sartori 1970: 1051). Decentralization, but to which level of governance? Knowing whether a state is more or less centralized tells one nothing about which tier does what. Decentralization measures focus on the central state, lumping together all levels of subnational governance as “the other,” the non-central state. This can be a useful simplification in cross-national comparison, but it severely restricts the study of governance within the state. It has nothing to say to cases where one level of regional governance is empowered at the expense of another. “How does one compare two three-tier systems, A and B, when in A one-third of the issues are assigned to each of the tiers, while in B 90 percent of the issues are assigned to the middle tier and 5 percent each to the top and bottom tiers” (Treisman 2007: 27; Oates 1972: 196). One needs to map individual regions and regional tiers to probe variation in multilevel governance.

The measure we propose builds on the concepts of federalism and decentralization (Enderlein et al. 2010; Oates 1972, 2005, 2006; Stein and Burkowitz 2010). Both ways of thinking about authority have been influential in our work, as in the discipline of political science as a whole. From federalism, our measure takes the idea that regional authority consists of distinct forms of rule: self-rule within a region and shared rule within the country as a whole. This provides us with the conceptual frame for our measure. From decentralization, the measure takes the idea that the structure of government can be measured along continuous variables that together summarize regional authority.

II. *The Specified Concept: Validity and Minimalism*

Our focus in this book is on legal authority which is

- *institutionalized*, i.e. codified in recognized rules;
- *circumscribed*, i.e. specifying who has authority over whom for what;
- *impersonal*, i.e. designating roles, not persons;
- *territorial*, i.e. exercised in territorially defined jurisdictions.

These characteristics distinguish legal authority from its traditional, charismatic, and religious variants. Weber (1968: 215–16) observes that “In the case

Transparency in Measurement

of legal authority, obedience is owed to the legally established impersonal order. It extends to the persons exercising the authority of office under it by virtue of the formal legality of their commands and only with the scope of authority of the office." The exercise of legal authority over a large population involves a minimum level of voluntary compliance with codified rules that have a specific sphere of competence, and which are exercised through formal institutions, including a differentiated administration (Weber 1968: 212–17).

A focus on legal authority has two benefits. The first is that it distinguishes the structure of government from causally related but conceptually distinct phenomena such as the organization of political parties, the ideological beliefs of those in office, or the incidence of corruption. The second is that legal authority can be evaluated using public records: constitutions, laws, executive orders, statutes, or other written documents which are publicly available to researchers who can confirm, revise, or refute our coding decisions.

Our approach is minimalist. Minimalism is a concept used in design to expose the essence of a form by eliminating all non-essential features. In measurement this is the effort to specify the essential properties of a concept by eliminating its superfluous connotations. This avoids entangling phenomena that one wishes to explore empirically. If a measure of subnational authority were to include an indicator for party centralization it would not help one investigate how party organization shapes the structure of government.

Minimalism and validity often exist in tension. Public spending might be considered a minimalist indicator of decentralization, but the proportion of public expenditure that passes through a subnational government does not tell us whether that government can determine spending priorities (see Chapter Two).

Where the rule of law is weak, informal practices may undercut provisions codified in law. Bertrand (2010: 163) summarizes the problem: "[A]utonomy can sometimes become an empty shell. Powers may exist in law, but are subsequently undermined by the central state. For instance, the central state can enact other legislation that might contradict the autonomy law. By various bureaucratic or extra-institutional means, it might also slow or stall the autonomy law's implementation. Repressive policies might be launched after the autonomy law is passed, thereby reducing its meaning and ultimately its legitimacy" (see also Eaton et al. 2010; Varshney, Tadjoeeddin, and Panggabean 2008). In many regimes, as O'Donnell (1998: 8) observes, "Huge gaps exist, both across their territory and in relation to various social categories, in the effectiveness of whatever we may agree that the rule of law means."

The measure we propose taps authority codified in law, but we do not interpret this mechanistically. Some written rules never make it into practice. If the constitution states that subnational governments may tax their own populations, yet enabling legislation is not enacted (as in *departamentos* and

provincias in Peru), then we do not consider the regions to have fiscal authority.¹⁷ Similarly, we code the date when a reform takes place, not when it is prescribed in legislation.¹⁸

We estimate reforms that are not enacted in law if they are codified in executive orders, decrees, or edicts that are considered legally binding. For example, we take into account the capacity of a central state to sack regional governors, as in Argentina under military rule, even though it had a flimsy legal basis. Article six of the Argentine constitution allows federal intervention only in a handful of circumstances such as civil war and violation of the constitution, but when a military junta came to power in 1966, it drafted a military decree, the *Acta de la Revolución*, which sanctioned centralization and the abrogation of civilian rule (Potash 1980: 195–6).

Eaton, Kaiser, and Smoke (2010: 24) point out that “complete institutional analysis must consider informal social norms that govern individual behavior and structure interaction between social actors.” This is true, but no measure should try to cover the entire field. To what extent should one include informal social norms in a measure of regional authority? This depends on the purpose of the measure. On the one hand, we wish to evaluate the concept of regional authority broadly to capture its reality, not just its appearance. On the other hand, we want to make it possible for researchers to investigate the causal links between the structure of government and its causes and consequences. If we included indicators for regime type, corruption, or clientelism in a measure of regional authority this would complicate causal inference.

For the same reason we leave partisanship and party politics aside. Regional governments may be more assertive if they have a different partisan complexion from that of the central government, but our focus is on the rules of the game rather than how they affect behavior. In Malaysia, for example, we code the capacity of Sabah and Sarawak to levy an additional sales tax without prior central state approval, even though this authority was used only from 2008 when opponents of the ruling *Barisan Nasional* coalition won regional

¹⁷ The 1933 and 1979 constitutions gave *departamentos* extensive fiscal authority with the capacity to set rate and base of certain taxes. However these provisions were not translated in enabling legislation, and a 1988 law mandating that national government would transfer property and income tax to the regions within three years was not implemented (Dickovick 2004: 7). The 1979 constitution also appeared to give *provincias* extensive fiscal authority, including property tax, vehicle tax, and construction tax (C 1979, Art. 257), but consecutive governments have interpreted these competences narrowly and continue to set the base of all taxes while imposing narrow bands for rates (Ahmad and García-Escribano 2006: 15; von Haldenwang 2010: 651).

¹⁸ The gap between legislation and implementation can be extensive. In South Korea it took twelve years for the Local Autonomy Act of 1988 to come into force. We code only the parts of the reform at the time they are implemented by enabling legislation (Bae 2007; Choi and Wright 2004). In Argentina, the 1994 constitution introduced direct elections for senators to replace appointment by the provincial legislature. The first direct elections took place in 2001, which is when we score direct election.

Transparency in Measurement

elections. If one is interested in finding out how political parties affect the exercise of authority, it makes sense to estimate political parties independently from the structure of government (Chhibber and Kollman 2004; Harbers 2010; Hopkin and Van Houten 2009; Riker 1964).

Regime variation poses a particular challenge given the expectation that dictatorship and centralization are related (Bird and Vaillancourt 1998; Elazar 1995; IADB 1997; Leff 1999). We want to pick up the effect of a regime in constraining or facilitating regional authority, but we do not want to build regime type into a measure of regional authority. One can expect authoritarianism to bias subnational relations toward centralization, but this is not a black-and-white phenomenon (Eaton 2006; Eaton et al. 2010; Gibson 2004; Montero and Samuels 2004; O'Neill 2005; Willis, Garman, and Haggard 1999). Authoritarian regimes typically suspend or abolish subnational legislatures or executives, but the extent, form, and timing varies considerably.

Some examples suggest the need for a nuanced approach. Whereas the *Revolución Argentina* (1966–72) replaced all elected governors and put provincial legislatures under military control, the coups in 1955 and 1964 left subnational institutions more or less intact (Eaton 2004a; Falleti 2010). The military regime in Brazil (1964–82) maintained direct elections for governorships for three years before requiring regional assemblies to select governors from a central list (Samuels and Abrucio 2000). Regional assembly elections were never canceled. Cuba's Castro regime sidelined provincial and municipal institutions in favor of sectoral juntas, but reintroduced them in 1966 (Roman 2003; Malinowitz 2006; Mendez Delgado and Lloret Feijoo 2007). In Indonesia, centralization under authoritarian rule was incremental. Provincial and municipal legislatures continued to be elected even under Suharto, and subnational executives were gradually brought under central control. In 1959, regional governors became dual appointees; in 1974, they were centrally appointed; and from 1979 the central government appointed mayors and district heads as well.

We also see some exceptional cases in which authoritarian rulers create a new regional level. In Chile, Pinochet created an upper level of fifteen deconcentrated *regiones* to empower his rural constituencies. He also shifted authority over schools and hospitals to municipal governments to weaken public sector unions. Both *regiones* and municipalities became focal points for subsequent decentralization (Eaton 2004c).

Regime change can have different effects for regional governance in different parts of a country. Democratization in Spain produced a cascade of regional bargains, beginning with the historic regions of the Basque Country, Catalonia, and Galicia. The 1978 constitution laid out two routes to regional autonomy, but competitive mobilization spurred a variety of institutional arrangements (Agranoff and Gallarín 1997).

A democratic opening is often followed by the accommodation of a previously suppressed ethnic minority. One result is that a country that had a homogenous structure of government becomes territorially differentiated. Aceh and Papua became autonomous Indonesian regions after Suharto's resignation (Bertrand 2007; Reid 2010b). Mindanao became an autonomous Philippine region following the People Power Revolution (Bertrand 2010: 178). Democratization in Russia after 1989 saw a series of bilateral arrangements with the central government empowering ethnic provinces (*respubliki*) (Svendsen 2002: 68–70).

A valid measure of regional authority should be sensitive to these phenomena. Theory in this rapidly growing field often engages the timing and character of regional authority, and it often has implications for individual regions as well as countries. If one wishes to test a theory relating democratization to multilevel governance, it is necessary to have measures in which these phenomena do not contaminate each other.

III. *Dimensions of Self-rule and Shared Rule*

One of the most important tasks in measuring an abstract concept is to decompose it into dimensions which a) can be re-aggregated to cover the meaning of the specified concept, b) are concrete in the sense that they are a step closer to observed reality, and c) are simple in that they are unidimensional and substantively interpretable (De Leeuw 2005). This can take more than one step. Measurement of the nominal GDP of the US begins by decomposing the concept into five categories—consumption, services, investment, exports, and imports—each of which is further disaggregated. Consumption, for example, consists of rental income, profits and proprietors' income, taxes on production and imports less subsidies, interest, miscellaneous payments, and depreciation. The purpose is to break down an abstract concept, in this case nominal GDP, into pieces that capture its content and can be empirically estimated (Landefeld et al. 2008). Similarly, measures of democracy disaggregate the concept into domains that can be broken down into dimensions (Coppedge et al. 2008, 2011).

Our first move is to distinguish two domains that encompass the concept of regional authority. *Self-rule* is the authority that a subnational government exercises in its own territory. *Shared rule* is the authority that a subnational government co-exercises in the country as a whole. The domains of self-rule and shared rule provide an elegant frame for our measure and they are widely familiar in the study of federalism (Elazar 1987; Keating 1998, 2001; Lane and Errson 1999; Riker 1964). The distinction appears to have empirical as well as theoretical bite. Research using our prior measure for OECD countries finds that self-rule and shared rule have distinct effects on corruption (Neudorfer

Transparency in Measurement

and Neudorfer 2015), spatial disparities (Ezcurra and Rodriguez-Pose 2013), regional representation (Donas and Beyers 2013; Tatham and Thau 2013), regional party vote share in national elections (Kyriacou and Morral-Palacin 2015), subnational coalition formation (Bäck et al. 2013), protest (Quaranta 2013), and voting (Niedzwiecki and Stoyan 2015).¹⁹

Self-rule and shared rule are distinct domains of regional governance. But we need to decompose them into dimensions to estimate variation.

The tripartite distinction between fiscal, administrative, and political decentralization is a useful point of departure. Fiscal decentralization is control over subnational revenue generation and spending; administrative decentralization is the authority of subnational governments to set goals and implement policies; and political decentralization refers to direct elections for subnational offices (Montero and Samuels 2004; Falleti 2005).²⁰ The four types of political decentralization identified by Treisman (2007: 23–7) overlap with this three-fold schema, with the important addition of a dimension for constitutional decentralization (“subnational governments or their representative have an explicit right to participate in central policy making”).

The revenue generating side of fiscal decentralization can be broken down into the authority of a regional government to control the base and rate of major and minor taxes and its latitude to borrow on financial markets without central government approval. On administrative decentralization it would be useful to know the extent to which the central government can veto subnational government and the kinds of policies over which subnational governments exert authority. And on political decentralization, one might distinguish between indirect and direct election of offices, and further, between the election of regional assemblies and regional executives.

Fiscal, administrative, and political decentralization are concerned with the authority of a regional government in its own jurisdiction. However, a regional government may also co-determine national policies. Is the regional government represented in a national legislature (normally the second chamber), and if so, to what effect? Can the regional government co-determine the proportion of national tax revenue that goes into its pocket? Does it have routinized access to extra-legislative channels to influence the national government? And, most importantly, does the regional government have authority over the rules of the game?

¹⁹ An incipient literature examines the diverse causes of self-rule and shared rule (see e.g. Amat and Falcó-Gimeno 2014). Joan-Josep Vallbe (2014) extends the self-rule/shared rule distinction to judicial regional authority.

²⁰ Falleti (2010: 329) takes a step toward a more specific conceptualization of administrative decentralization as “the set of policies that transfer the administration and delivery of social services such as education, health, social welfare, or housing to subnational governments.”

Measurement

These distinctions provide a basis for further specification. Each responds to a basic question that one can ask about regional authority. In the domain of *self-rule* we formulate five questions:

- How independent is a regional government from central state control? *Institutional depth* tracks the extent to which a regional government can make autonomous policy decisions. A deconcentrated regional administration has the apparatus of government—a physical address, a bureaucracy, an executive, a budget—but is subordinate to the center. A decentralized regional government, by contrast, can make independent policy decisions, which, at the upper end of this scale, are not subject to central government veto.
- What is the range of a regional government's authority over policy within its jurisdiction? *Policy scope* taps the breadth of regional self-rule over policing, over its own institutional set-up, over local governments within its jurisdiction, whether a regional government has residual powers, and whether its competences extend to economic policy, cultural-educational policy, welfare policy, immigration, or citizenship.
- What authority does a regional government have over taxation within its jurisdiction? *Fiscal autonomy* is evaluated in terms of a regional government's authority to set the base and rate of minor and major taxes in its jurisdiction. This dimension is concerned with the authority of a government to set the rules for taxation rather than the level of regional spending.
- Does a regional government have authority to borrow on financial markets? *Borrowing autonomy* evaluates the centrally imposed restrictions on the capacity of a regional government to independently contract loans on domestic or international financial markets.²¹
- Is a regional government endowed with representative institutions? *Representation* assesses whether a regional government has a regionally elected legislature; whether that legislature is directly or indirectly elected; and whether the region's executive is appointed by the central government, dual (i.e. co-appointed by the central government), or autonomously elected (either by the citizens or by the regional assembly).

²¹ Our prior measure overlooked borrowing (Hooghe, Marks, and Schakel 2008, 2010). Extending the sample to Latin America and South-East Asia brings regional borrowing into focus both in self-rule and shared rule. Subnational borrowing became particularly salient from the 1980s and 1990s when several Latin American countries were hit by debt crises. The financial crisis in the Eurozone has also put the spotlight on regional borrowing.

Transparency in Measurement

In the domain of *shared rule* we pose the following questions:

- To what extent can a regional government co-determine national policy making? *Law making* assesses the role of regions in structuring representation at the national level (i.e. in a second legislative chamber); whether regions have majority or minority representation there; and the legislative scope of the second chamber.
- Can a regional government co-determine national executive policy in intergovernmental fora? *Executive control* taps whether regional governments have routine meetings with the central government and whether these are advisory or have veto power.
- Can a regional government co-determine how national tax revenues are distributed? *Fiscal control* taps the role of regions in negotiating or exerting a veto over the territorial allocation of national tax revenues.
- Can a regional government co-determine the restrictions placed on borrowing? *Borrowing control* distinguishes whether regional governments have no role, an advisory role, or a veto over the rules that permit borrowing.
- Can a regional government initiate or constrain constitutional reform? *Constitutional reform* assesses the authority of a regional government to propose, postpone, or block changes in the rules of the game. Does constitutional reform have to gain the assent of regional governments or their constituencies? Does it require majority support in a regionally dominated second chamber?

A region may exercise shared rule multilaterally with other regions or it may exercise shared rule bilaterally with the center. Multilateral shared rule is contingent on coordination with other regions in the same tier; bilateral shared rule can be exercised by a region acting alone (Chapter Three).

IV. Indicators for Dimensions of Self-rule and Shared Rule

An indicator consists of rules for inferring variation along a dimension (Tal 2013: 1162; King, Keohane, and Verba 1994: 75). Chang (2004: 216) asks, “In the process of operationalizing the abstract concept, what exactly do we aim for, and what exactly do we get? The hoped-for outcome is an agreement between the concrete image of the abstract concept and the actual operations that we adopt for an empirical engagement with the concept (including its measurement).”

Our purpose is to devise indicators that encompass the meaning of the concept and can be reliably scored. All observations, even simple ones like the number of votes received by a candidate in an election, are contestable,

but some observations are more contestable than others (Lakatos 1970). For example, an indicator that asks a coder to score “the ability of the center to suspend lower levels of government or to override their decisions” (Arzaghi and Henderson 2005) is abstract and ambiguous.²² What if there are several lower levels of government and they differ? What if the central government can suspend a lower level government only under exceptional circumstances? What if some lower level decisions may be overridden and others not?

Tables 1.2 and 1.3 detail indicators for self-rule and shared rule.²³ The indicators specify institutional outcomes for an individual region or regional tier that can be reliably assessed against information in constitutions, laws, executive orders, government documents. In addition, the intervals are designed to have the following desiderata (Gerring and Skaaning 2013; Goertz 2006):

- Each interval is comprised of a set of necessary and sufficient institutional conditions for a particular score.
- The attributes for each interval encompass the prior interval with some additional unique attribute.
- The attributes are binary in order to minimize the gray zone between existence and non-existence.
- Collectively, the intervals seek to capture the relevant variation in the population that is assessed.
- The spacing of the intervals is conceived as equidistant so that a unit shift along any dimension is equivalent.

V. Scoring Cases

Scoring cases consists of obtaining and processing information in order to place numerical values on objects (Bollen and Paxton 2000). Our scoring strategy involves “interpretation through dialogue.”

Interpretation is the act of explaining meaning among contexts or persons. When measuring regional authority we are interpreting the concept of regional authority in the context of particular regions at particular points in time. As one moves down the ladder of measurement in Figure 1.1, the

²² “This dimension measures whether or not the central government has the legal right to override the decisions and policies of lower levels of government. If the central government has such a right, the country scores zero; if not, the score is four. To ‘override’ in this context means to be able to veto without due process. Many countries have legal mechanisms for the appeal and review by higher authorities of lower-level government decisions. As a rule, these do not constitute override authority, unless they are extremely lax. Instead, override authority exists when the central government can legally deny regional and local authority with an ease that calls that very authority in to question.” <<http://www.econ.brown.edu/faculty/henderson/decentralization.pdf>>.

²³ Law making consists of four sub-dimensions.

Table 1.2. Self-rule

Self-rule	The authority exercised by a regional government over those who live in the region
Institutional depth	<p>The extent to which a regional government is autonomous rather than deconcentrated.</p> <p>0–3</p> <p>0 No functioning general purpose administration at regional level. 1 Deconcentrated, general purpose, administration. 2 Non-deconcentrated, general purpose, administration subject to central government veto. 3 Non-deconcentrated, general purpose, administration not subject to central government veto.</p>
Policy scope	<p>The range of policies for which a regional government is responsible.</p> <p>0–4</p> <p>0 Very weak authoritative competencies in a), b), c), or d) whereby a) economic policy; b) cultural–educational policy; c) welfare policy; d) one of the following: residual powers, police, own institutional set-up, local government. 1 Authoritative competencies in one of a), b), c), or d). 2 Authoritative competencies in at least two of a), b), c), or d). 3 Authoritative competencies in d) and at least two of a), b), or c). 4 Criteria for 3 plus authority over immigration, citizenship, right of domicile.</p>
Fiscal autonomy	<p>The extent to which a regional government can independently tax its population.</p> <p>0–4</p> <p>0 Central government sets the base and rate of all regional taxes. 1 Regional government sets the rate of minor taxes. 2 Regional government sets the base and rate of minor taxes. 3 Regional government sets the rate of at least one major tax: personal income, corporate, value added, or sales tax. 4 Regional government sets base and rate of at least one major tax.</p>
Borrowing autonomy	<p>The extent to which a regional government can borrow.</p> <p>0–3</p> <p>0 The regional government does not borrow (e.g. centrally imposed rules prohibit borrowing). 1 The regional government may borrow under prior authorization (<i>ex ante</i>) by the central government and with one or more of the following centrally imposed restrictions a. golden rule (e.g. no borrowing to cover current account deficits) b. no foreign borrowing or borrowing from the central bank c. no borrowing above a ceiling d. borrowing is limited to specific purposes. 2 The regional government may borrow without prior authorization and under one or more of a), b), c), or d). 3 The regional government may borrow without centrally imposed restrictions.</p>
Representation	<p>The extent to which a region has an independent legislature and executive.</p> <p>0–4</p> <p>Assembly: 0 No regional assembly. 1 Indirectly elected regional assembly. 2 Directly elected assembly. Executive: 0 Regional executive appointed by central government. 1 Dual executive appointed by central government and regional assembly. 2 Regional executive is appointed by a regional assembly or directly elected.</p>

Table 1.3. Shared rule

Shared rule		The authority exercised by a regional government or its representatives in the country as a whole	
Law making	The extent to which regional representatives co-determine national legislation.	0–2	0.5
		0.5	0.5
Executive control	The extent to which a regional government co-determines national policy in intergovernmental meetings.	0–2	0
		1	1
Fiscal control	The extent to which regional representatives co-determine the distribution of national tax revenues.	0–2	0
		1	1
Borrowing control	The extent to which a regional government co-determines subnational and national borrowing constraints.	0–2	0
		1	1
Constitutional reform	The extent to which regional representatives co-determine constitutional change.	0–4	0
		1	1
		2	2
		3	3
		4	4

Regions are the unit of representation in a national legislature. Regional governments designate representatives in a national legislature. Regions have majority representation in a national legislature based on regional representation.

The legislature based on regional representation has extensive legislative authority.

No routine meetings between central and regional governments to negotiate policy. Routine meetings between central and regional governments *without* legally binding authority.

Routine meetings between central and regional governments *with* legally binding authority.

Neither the regional governments nor their representatives in a national legislature are consulted over the distribution of national tax revenues.

Regional governments or their representatives in a national legislature negotiate over the distribution of tax revenues, but do not have a veto.

Regional governments or their representatives in a national legislature have a veto over the distribution of tax revenues.

Regional governments are not routinely consulted over borrowing constraints.

Regional governments negotiate routinely over borrowing constraints but do not have a veto.

Regional governments negotiate routinely over borrowing constraints and have a veto.

The central government or national electorate can unilaterally reform the constitution. A national legislature based on regional representation can propose or postpone constitutional reform, raise the decision hurdle in the other chamber, require a second vote in the other chamber, or require a popular referendum.

Regional governments or their representatives in a national legislature propose or postpone constitutional reform, raise the decision hurdle in the other chamber, require a second vote in the other chamber, or require a popular referendum.

A legislature based on regional representation can veto constitutional change; or constitutional change requires a referendum based on the principle of equal regional representation.

Regional governments or their representatives in a national legislature can veto constitutional change.

Transparency in Measurement

concept of regional authority becomes less abstract, but even concrete concepts, such as a dual executive, a routine meeting, or a formal veto, are not directly observable. “The bridge we build through acts of measurement between concepts and observations may be longer or shorter, more or less solid. Yet a bridge it remains” (Schedler 2012: 22). Our intent is to make the link between indicators and scores both plausible and transparent.

Dialogue—sustained, open-ended discussion—is intended to increase the validity of our judgments. While time intensive, dialogue among coders is vital for consistent interpretations across countries. Bowman, Lehoucq, and Mahoney (2005: 957) describe the process which underpins their democracy index as iterative consensus building: “Disagreements arose regarding the codes for several particular measures, and these differences generally reflected either a limitation in the measure or a limitation in an author’s knowledge of the facts. If the problem was with the resolving power of a measure, we sought to better define the measure until a consensus could be reached. If the problem arose not because of the measure but rather because of divergent understandings of the empirical facts, we reviewed all evidence and argued about the facts.” Our approach is similar (see also Saylor 2013).

Dialogue among coders makes it impossible to assess inter-coder reliability, but this is a sacrifice worth making. The principal challenge in estimating an abstract concept such as regional authority is validity rather than reliability. Validity concerns whether a score measures what it is intended to measure. Do the dimensions really capture the meaning of the concept? Do the indicators meaningfully pick up the variation on each dimension? Do the scores accurately translate the characteristics of individual cases into numbers that express the underlying concept? Reliability concerns the random error that arises in any measurement. How consistent are scores across repeated measurements? Would a second, third, or *n*th expert produce the same scores? If the error one is most worried about is systematic rather than random, then it may be more effective to structure dialogue among coders to reach consensus on a score than to combine the scores of independent coders.

Using expert evaluations is inappropriate for the data we seek. Expert surveys are useful for topics that are “in the head” of respondents. The information required to assess the authority of individual regions in a country on ten dimensions annually from 1950 goes far beyond this. It is not a matter of providing proper instructions to experts. The limitations of expert surveys are more fundamental (Steenbergen and Marks 2007; Marks et al. 2007).²⁴

²⁴ Expert surveys are an economical and flexible research tool when the information necessary for valid scoring is directly accessible to the experts (Wiesehomeier and Benoit 2009). The number of experts need not be large—a rule of thumb would be six or more for each observation (Steenbergen and Marks 2007; Marks et al. 2007). Expert surveys eliminate the need to have specific sources of information (e.g. laws, government documents) available for all cases. And

Measurement

An evaluation based on a series of expert surveys over fifteen years concludes that “Unambiguous question wording is necessary but not sufficient for reliable expert judgments. Perhaps the most important source of error lies neither in poor question-wording, nor in the selection of experts, but in asking questions that lie beyond the expertise of respondents” (Hooghe et al. 2010: 692). This limitation, along with our overriding concern with validity, suggests that dialogue among researchers is both more feasible and more appropriate than an expert survey for the task at hand.

The practical steps involved in interpretation through dialogue are as follows:

- *Gathering and interpreting public documents.* An initial step is to collect publicly available information related to the indicators. These are first and foremost constitutions, laws, executive decrees, budgets, government reports, and websites.²⁵ This is usually not so difficult for the most recent one or two decades, but can be challenging for the 1950s and 1960s.
- *Engaging the secondary literature.* Numerous books, articles, and non-governmental studies cover the larger and richer countries. The coverage of Latin America, Eastern Europe, and Asia has increased markedly in recent years. However, secondary sources thin as one goes back in time.²⁶ In most cases, the secondary literature is less useful as a source of “facts” than it is as a conceptual/theoretical basis for probing our measurement decisions, including particularly the contextual appropriateness of the indicators.
- *Subjecting interpretations to expert commentary.* Although it is unreasonable to expect country experts to provide strictly comparable scores for individual regions across ten dimensions on an annual basis going back to 1950, they can provide valuable feedback on the validity of scoring judgments. For countries that we regard as the most complex or least sourced, we commissioned researchers who have published extensively on

expert surveys are flexible tools for experiments designed to evaluate and improve the reliability of the measure. It is possible to introduce vignettes into the survey that tell us how individual experts evaluate benchmark scenarios (Bakker et al. 2014). However, the virtues of expert surveys are null if experts are asked to evaluate topics to which they do not have direct cognitive access. In the Chapel Hill Expert Survey, we have found that items tapping expert judgments on the contemporary positioning of political parties on major issues produce reliable scores, while items that ask experts for more specific information on the extent of division within political parties on those same issues fail to do so. The information that we seek on regional authority is much more specific than that required for evaluating divisions within political parties.

²⁵ Wikipedia lists territorial subdivisions for most countries, and <<http://www.statoids.com>>, a website run by Gwillim Law, a Chapel Hillian, is a fount of information.

²⁶ Country reports from the OECD’s multilevel governance unit are valuable sources. Also useful are studies commissioned by the Inter-American Development Bank, the Asian Development Bank, and the World Bank.

Transparency in Measurement

regional authority in their country to write commentaries on our interpretations and scores. These commentaries led us back to the primary sources, and in some cases to revisit our conceptualization of the indicators and the dimensions.

- *Discussing contending interpretations in extended dialogue.* All scoring decisions were discussed by three or more members of the research team, often at length. Difficult cases were usually discussed on more than two occasions. Divergence of interpretation led us to soak and poke by going back to the sources or finding additional sources. It was also instrumental in refining the indicators, and led us to distinguish between bilateral and multilateral shared rule.²⁷ Interpretation through dialogue made it possible to revisit our decisions on indicators and dimensions as we sought to place institutional alternatives in diverse countries on a single theoretical–conceptual frame.
- *Paying sustained attention to ambiguous and gray cases.* No matter how well designed a measure, there will always be ambiguities in applying rules to particular cases. There will also be gray cases that lie between the intervals. Our approach is to clarify the basis of judgment and, where necessary, devise additional rules for adjudicating such cases that are consistent with the conceptual underpinnings of the measure. Chapter Three sets out our rules for coding ambiguous and gray cases and is, not coincidentally, the longest chapter in this book.
- *Explicating judgments in extended profiles.* The lynchpin of our measure is the endeavor to explain coding decisions. This involves disciplined comparison across time and space. The country profiles in this volume make our scoring evaluations explicit so that researchers familiar with individual cases may revise or reject our decisions. At the same time, the profiles are intended to remove the curtain that protects the cells in a dataset from cross-examination.

VI. Adjudicating Scores

Gray cases are endemic in measurement. They come into play at every step in a measure and arise in the fundamental tension, noted by Weber, between an idea and an empirical phenomenon. Gray cases are not indicators of scientific failure. Rather they are calls for re-assessing a measurement, for ascending the arrows on the right side of Figure 1.1. One can seek to resolve a gray case by refining observation, by revising an indicator, dimension or, *in*

²⁷ See the appendix for the coding schema for multilateral and bilateral shared rule.

extremis, by redefining the specified concept. Is this case gray because we lack good information or does it raise conceptual issues? Is the case an isolated instance of ambiguity or does it suggest a more general problem? If the latter, can one rejig the indicator for that dimension? Or does the problem go back to the specification of the concept?

Gray cases contain valuable information for users and for those who might wish to improve a measure. They flag areas for improving a measure. We notate three common sources of “grayness” in the extensive country profiles in Part II.

- *Insufficient or ambiguous information.* Outside the laboratory, observation can be plagued by poor light or deficient information. We indicate scores for which we have thin information by using the symbol α in superscript in the profile.
- *Observations that fall in-between intervals.* No matter how sharp a distinction, some observations sit between intervals. We indicate these borderline cases with the symbol β in superscript.
- *Disagreement among sources, coders, experts.* Applying a concept to an empirical phenomenon is an inferential process that is subject to error and hence to disagreement. Even simple concepts that refer to physical objects have fuzzy boundaries (Quine 1960: 114ff). We note disagreements among sources, coders, and/or experts with the superscript γ .

Conclusion

Measuring the authority of individual regions in a wide range of countries over several decades is always going to be a theoretical as well as practical challenge. Our approach, in short, is to a) disaggregate the concept into coherent dimensions that encompass its meaning; b) operationalize these dimensions as institutional alternatives that are abstract enough to travel across cases but specific enough to be reliably evaluated; c) assess the widest possible range of documentary information in the light of the secondary literature and expert feedback; and d) discuss coding decisions and ambiguities in comprehensive country profiles.

The measure can be used to estimate regional authority at the level of the individual region, regional tier, or country by combining the dimensions. Alternatively, researchers may wish to re-aggregate these to their needs. The intervals on the dimensions are conceptualized along equal increments, so one can sum dimension scores to produce a scale ranging between 1 and 30 for each region or regional tier. Country scores are zero for countries that have no regional government, but there is no a priori maximum because countries may

Transparency in Measurement

Table 1.4. Polychoric factor analysis

Components	Single-factor solution	Two-factor solution:	
		Self-rule	Shared rule
Institutional depth	.86	.87	.08
Policy scope	.91	.88	.13
Fiscal autonomy	.84	.59	.34
Borrowing autonomy	.85	.86	.08
Representation	.81	.99	–.12
Law making	.74	.08	.76
Executive control	.82	.12	.80
Fiscal control	.75	.04	.81
Borrowing control	.62	–.08	.77
Constitutional reform	.78	.05	.83
Eigenvalue	6.43	5.51	5.29
Chi-squared	859.38	859.38	
Explained variance (%)	81.9		
Factor correlation		0.61	

Note: Principal components factor analysis, promax non-orthogonal rotation, listwise deletion. $n = 80$ (country scores in 2010). For the two-factor solution, the highest score for each dimension is in bold.

have more than one tier. Chapter Three explains how we aggregate regional scores into country scores. We use this additive scale in the maps, tables, and figures in this book.

An alternative approach is to interpret the dimensions as indicators of a latent variable. The Cronbach's alpha across the ten dimensions for 2010 is 0.94, which suggests that the dimensions can be interpreted as indicators of a single latent concept. Table 1.4 presents a factor analysis for country scores in 2010. We use polychoric correlations on the conservative assumption that the indicators are ordinal. A single-factor solution accounts for 82 percent of the variance. When we impose a two-factor solution, each indicator loads strongly on one latent factor and weakly on the other factor. The solution confirms the theoretical distinction between self-rule and shared rule.²⁸

It does not make much difference which method one uses to aggregate the data. The scores derived from factor analysis and from additive scaling are very similar. The correlation is 0.98 for 2010 for the single dimension. Figure 1.2 plots correlations using interval data and shows that the index is robust across alternative weights for self-rule and shared rule. The RAI weighs shared rule to self-rule in the ratio of 2:3. When we reverse these weights, the rank order among countries in 2010 yields a Spearman's *rho* of 0.99 (Pearson's $r=0.97$).

The decision to estimate authority at the level of individual regions rather than countries is the single most important decision in this book because it affects how one thinks about the structure of governance. Governance

²⁸ The correlation between the two dimensions is reasonably strong ($r=0.61$).

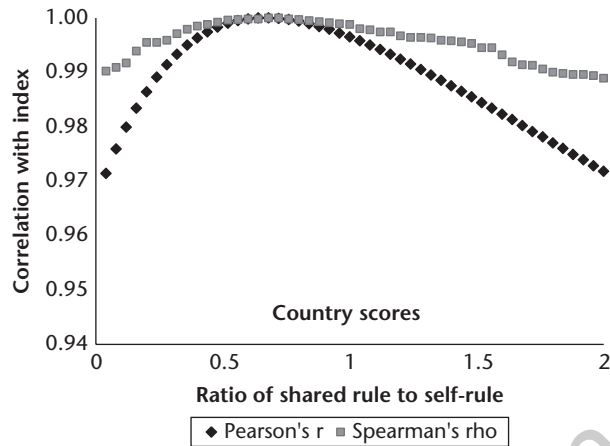


Figure 1.2. Robustness of the regional authority index across alternative weights for self-rule and shared rule

Note: Calculations are for 2010; $n = 80$. Spearman's *rho* is calculated on ordinal scores, and Pearson's *r* is calculated on the interval scores. The RAI weights shared rule to self-rule in the ratio of 2:3 (0.66). Here we vary the ratio between 0 and 2.

exhibits great variation within as well as among countries, and one cannot begin to fathom the reasons for this or understand its consequences if one conceives the state as the unit of analysis. Some regional governments have wide ranging policy competences; others deal with a single problem. Some can block constitutional reform. Some have extensive taxing powers. Some exert wide ranging authority within their own territories; others play a decisive role in the governance of the country as a whole. Some regions have a special bilateral relationship with the central government, while others exist alongside other regions in uniform tiers. The variation that the RAI detects among countries is extremely wide, and now one can also systematically probe variation within countries over time.

Finally, the effort to measure a concept as complex as regional authority may have implications for measurement in general. Measurement seeks to establish a numerical relation between an observable phenomenon and a concept. This, as Max Weber emphasized, involves interpretation. What, precisely, is being measured? How is the concept specified? What are its dimensions? How are intervals along these dimensions operationalized? How are individual cases scored on those dimensions? What rules apply to gray cases? These are questions that confront social science measurement generally. Each question involves judgment, the weighing of one course of action against others. Our goal in this book is to make those judgments explicit, and hence open to disconfirmation or improvement.