2

Crossvalidating the Regional Authority Index

Our aim in this chapter is to assess the validity of the regional authority index (RAI) by comparing it to prior institutional and fiscal measures. We begin by asking whether alternative institutional measures give similar scores to the same cases. This is convergent validation, the extent to which measures of the same concept are positively associated with each other (Bollen 1989: 188; Ray 2007: 12). To assess convergent validity we evaluate the extent to which these measures are in agreement with the RAI, explore sources of disagreement in a regression analysis, and complement this with an in-depth look at particular cases.

Convergence provides confidence in the validity of our measurement whereas disagreement provides a basis for further investigation. Each measure suffers from error, and the sources of error may vary in non-random ways. We find that differences among decentralization measures have systemic causes, both with regard to the extent of difference and the direction of difference. The most important differences arise because some countries have more than one tier of regional government between the local and the national and because measures seek to estimate decentralization over a period in which there has been extensive change.

Beyond such systematic differences, institutional measures sometimes arrive at sharply contrasting scores for individual countries, and the reasons for this are worth investigating in some detail. Knowing when, where, and how error in measurement arises helps one decide whether to use one measure over another (Adcock and Collier 2001; Bollen 1989; King, Keohane, and Verba 1994; Marks et al. 2007).

We conclude by discussing the content validity of three types of fiscal indicators and comparing their scores to the RAI. Content validity "assesses the degree to which an indicator represents the universe of content entailed in the systematized concept being measured" (Adcock and Collier 2001: 537).[1]

[1] Adcock and Collier (2001: 537) also identify a third type—criterion validity, which assesses "whether the scores produced by an indicator are empirically associated with scores for other

Here the task is to clarify the domain of the concept and to judge whether the measures fully represent the intended domain (Bollen 1989: 185). Are crucial elements omitted or are inappropriate elements included? Fiscal measures have the virtue of reliability, but we suggest that they do, indeed, omit important dimensions of decentralization and are correspondingly limited as a measure of decentralization.

Institutional Indicators of Decentralization

There is no shortage of measures of decentralization with which the RAI can be compared. Table 2.1 overviews the five most commonly used measures that, like the RAI, focus on the authoritative competences of subnational governments. All five measures conceive decentralization as a latent variable with fiscal, political, and administrative indicators (Falleti 2010; Schneider 2003). Each covers an array of countries on multiple dimensions of decentralization that can be summarized at the level of the country as a whole.

The chief differences between these measures and the RAI are as follows (Table 2.2):[2]

- *Unit of measurement.* The RAI is distinct in conceiving the individual region and the regional tier, rather than the country, as units of analysis. This increases the number of observations and makes it possible to compare regions and regional tiers within, as well as across, countries. We use aggregate RAI country scores for the purpose of comparison, but it is worth keeping in mind that country scores are just a useful fiction. The actual units of subnational authority in all decentralization measures are individual general purpose governments within territorially circumscribed jurisdictions.
- *Time period.* The RAI provides annual observations for 1950–2010. Brancati (2006, 2008) provides annual observations for 1985–2000. Arzaghi and Henderson (2005) assess eight five-year intervals between 1960 and 1995. Treisman (2002) is a cross-sectional measure,[3] and Lijphart (1999) and Woldendorp, Keman, and Budge (2000) average decentralization over several decades on the assumption that decentralization is fairly stable over time (Inman 2008).

variables, called criterion variables, which are considered direct measures of the phenomenon of concern." We do not assess criterion validity because there is no generally accepted "criterion variable" or "gold standard" for measuring regional authority.

[2] For descriptive statistics see Tables 2.A.1 and 2.A.2 in the appendix.

[3] Treisman (2002) is available as an unpublished paper on his website.

Crossvalidating the Regional Authority Index

Table 2.1. Institutional measures of decentralization

Measure	Dimensions	Indicators
Arzaghi and Henderson (2005)	Index of institutional decentralization, or effective federalism, consisting of: • formal government structure • political responsibilities of subnational governments • fiscal responsibilities of subnational governments	Effective federalism (0–4) is the average of: – constitutional federal versus unitary structure (0 or 4) – election of a regional executive (0 or 4) – election of a local executive (0 or 4) – ability of the center to suspend lower levels of government or to override their decisions (0 or 4) – revenue sharing (0, 2, or 4)
Brancati (2008)	Level of political decentralization: • elective dimension • policy dimension	Political decentralization (0–5) is the sum of five dichotomous indicators: – democratically elected regional legislatures – regional legislatures can raise or levy their own taxes – regional legislatures have joint or exclusive control over education – regional legislatures have joint or exclusive control over public order or police – regions must approve constitutional amendments
Lijphart (1999)	Federalism whereby countries are categorized on the basis of: • formal character of government structure (federal or unitary) • extent of decentralization (range of powers assigned to the regional level)	Federalism (1–5) is an ordinal scale: – unitary and centralized (=1) – unitary and decentralized (=2) – semi-federal (=3) – federal and centralized (=4) – federal and decentralized (=5)
Treisman (2002)	Decision making decentralization defined as formal rules about the distribution of political authority over decision making	Decision making decentralization (0–3) is an additive scale: – autonomy = the constitution reserves to subnational legislatures the exclusive right to legislate in at least one policy area – residual authority = the constitution assigns to subnational legislatures the exclusive right to legislate on issues that are not specifically assigned to one level of government; – subnational veto = a regionally elected upper chamber exists with the constitutional right to block legislation

Woldendorp, Keman, and Budge (2000)	Autonomy index consists of:	Autonomy index (0–8) is an additive scale:
	• fiscal centralization	– 2 if a country has a degree of fiscal centralization lower than 75%; – 1 if a country has fiscal centralization between 75% and 90%; – 0 if a country has fiscal centralization equal to or more than 90%
	• regional autonomy	– 2 if regional autonomy is formally laid down (federal states); – 1 if the country is a semi-federalist system; – 0 if neither
	• local government autonomy	– 2 if local government is mentioned in the constitution, its autonomy is recognized, and it is guaranteed direct representation; – 1 if one or two of these conditions are met; – 0 in all other cases
	• centralization	– 2 if the state is not centralized; – 1 if the state is medium centralized; – 0 if the state is highly centralized

Note: The operationalization of fiscal centralization diverges somewhat from the one published in Woldendorp, Keman, and Budge (2000). The adjustments were made after communication with Hans Keman and Jaap Woldendorp.

- *Tiers*. The RAI estimates the authority of subnational governments at each level between the local (>150,000 population) and the national. Arzaghi and Henderson and Woldendorp et al. are chiefly concerned with regional government, but have some items that encompass local government.[4] The remaining measures do not discriminate levels of government.
- *Dimensions*. All measures conceive decentralization as multidimensional. Arzaghi and Henderson, Brancati, Treisman, and the RAI estimate regional assemblies. Brancati, Woldendorp, and the RAI estimate regional tax authority. Treisman and the RAI evaluate whether residual powers rest with the region or the central state. In addition, the RAI estimates shared rule, the authority co-exercised by a region and regional tier within the country on five dimensions for law making, executive control, fiscal decision making, borrowing, and constitutional reform (Table 1.3).

[4] For this reason, in the following analyses we exclude the "election of a local executive" dimension from the Arzaghi and Henderson measure (see Table 2.1). We thank Christine Kearney for providing us with disaggregated scores. We are unable to exclude scores for local government in the Woldendorp et al. measure because disaggregated estimates are not available.

Table 2.2. Comparing decentralization measures

	Coverage			Concept specification			Concept disaggregation		Unit of measurement
	number of countries	time period	time points	multiple regional tiers	local/ regional separated	local tier included	number of dimensions	number of intervals	
Regional Authority Index	81	1950–2010	61	Yes	Yes	No	10	42	Region
Arzaghi/Henderson	48	1960–1995	8	No	Partial	Yes	6	14	Country
Brancati[a]	37	1985–2000	16	No	No	No	5	5	Country
Lijphart	36	1945–1996	1	No	No	Yes	2	5	Country
Treisman[b]	166	1990–1999	1	No	No	Yes	3	3	Country
Woldendorp et al.	51	1945–1998	1	No	Partial	Yes	4	8	Country

[a] The Brancati measure has twenty-three countries in common with the RAI, but the analysis reported here uses fifty-eight countries: the twenty-three countries coded by Brancati and an additional thirty-five coded by the authors on the basis of information provided by Dawn Brancati in personal communication. The smaller-N analysis is available from the authors upon request.
[b] It should be noted that Treisman is reluctant to aggregate the three indicators of decision making decentralization into a single index (see e.g. 2002: 9–10).

- *Country coverage*. Each measure covers the larger Western democracies (Table 2.A.1). Treisman covers virtually every non-micro state. The RAI covers all members of the Organisation of Economic Cooperation and Development (OECD), all Latin American countries, ten countries in Europe beyond the European Union (EU), and eleven in the Pacific and South-East Asia. Woldendorp et al. cover fifty-one democracies. Arzaghi and Henderson cover forty-eight countries with a population over ten million. Brancati covers thirty-seven countries with regional ethnic groups and Lijphart covers thirty-six democracies.
- *Intervals*. All measures go beyond the classic federal/unitary dichotomy. However, the number of intervals varies from three (Treisman) and five (Brancati; Lijphart) to forty-two (RAI) (see Table 2.2). The more fine-grained a measure, the better equipped it is to differentiate levels of decentralization among federal and among unitary countries. Lijphart's measure compresses nearly all federal countries at the high end of the scale with a score of five. Treisman's measure separates federal countries from each other but compresses most unitary non-federal countries in the lowest category.

Cross-sectional Comparison

To what extent do the measures tap a common dimension? Our first step is to conduct principal factor analyses on a cross-sectional dataset containing average country scores over time produced by each measure.[5] Since the country overlap varies across the measures, we conduct four factor analyses in the columns labeled "Country scores" in Table 2.3. We then use all the available data and conduct principal factor analyses for the same measures with annual observations for each country. These are the results displayed under "Country/year scores" in Table 2.3.[6]

The results reveal a high degree of convergence. In no comparison does the eigenvalue of the principal axis fall below 2.4, and the common variance is around 80 percent across the board. Every decentralization index loads heavily on the principal axis with factor loadings in excess of 0.74 for both cross-sectional and panel datasets. Notwithstanding the differences among the measures noted above, they appear to tap a common latent variable.

[5] Table 2.A.3 reports Pearson correlations.
[6] Countries without a region or regional tier are excluded in all analyses.

Table 2.3. Factor analysis of decentralization measures

	Country scores				Country/year scores			
	I	II	III	IV	I	II	III	IV
Regional Authority Index	0.96	0.93	0.93	0.92	0.93	0.89	0.92	0.92
Arzaghi/Henderson	0.85	–	–	–	0.91	–	–	–
Brancati	0.90	0.92	0.92	0.89	0.93	0.93	0.92	0.89
Lijphart	0.92	0.92	–	–	0.88	0.91	–	–
Treisman	0.74	0.88	0.91	0.91	0.80	0.90	0.92	0.92
Woldendorp et al.	0.96	0.87	0.87	–	0.94	0.85	0.86	–
N	10	21	31	58	70	148	265	558
Eigenvalue	4.76	4.09	3.29	2.47	4.83	4.01	3.26	2.47
Explained Variance (%)	79	82	82	82	81	80	81	82

Sources of Disagreement

High factor scores can hide significant differences in scoring that may have systematic sources (Marks et al. 2007). In this section, we consider several possible sources of disagreement among decentralization measures:

Limited Country Coverage

One might expect less researched countries to generate more disagreement than the "normal suspects," which in this field are the larger Western democracies. All six measures encompass a set of ten democracies in North America and Western Europe, but coverage declines as one moves to Eastern Europe, Southern Europe, South America, and Central America and the Caribbean. Table 2.A.1 in the appendix lists the countries covered by each measure. *Limited coverage* is the total number of times a country is excluded by the five alternative measures on the expectation that this will be positively associated with disagreement in scoring.

Distance in Time

Measurement error is likely to increase with retrospective evaluation. This is a particular problem for the RAI, which scores regions going back to 1950. The remaining measures have shorter time periods or provide single scores for multiple decades. We expect disagreement with the RAI to be higher for earlier than for later time periods both because the availability of information declines as one goes back in time and because time invariant measures may be biased toward recent years. The variable *Distance in time* is 2010 minus the year in which a country score is assessed.

Multiple Regional Tiers

The existence of multiple regional tiers in a country can produce different scores for measures that summarize all tiers or just the most authoritative tier. The RAI aggregates scores for all tiers between the local and the national, whereas the remaining measures do not explicitly distinguish different levels of subnational governance. *Tiers* is the number of regional tiers in a country. When a tier covers only part of a country, we weight each tier by the proportion of a country's population it encompasses.

Differentiation

Regions that have special authoritative competences that differentiate them from other regions in a country may give rise to scoring differences. Whereas the RAI estimates such regions individually and then aggregates regional scores to the country level using population weights, the remaining measures are national in focus. Differentiated governance is quite common: in 2010, thirty-five countries of the sixty-two countries included here had asymmetric, autonomous, or dependent regions (Hooghe and Marks forthcoming). *Differentiation* is calculated as the difference between the maximum and minimum RAI for units within the most authoritative regional tier.

Reform

The creation or abolition of regional tiers and reform in the authority of established regions may lead to scoring differences between measures that average decentralization over multiple years and those that have annual estimates. Our expectation is that disagreement will be greatest for countries where contemporary estimates provide weak guidance in estimating prior levels of decentralization.[7] The RAI detects jurisdictional reform in sixty countries that have one or more regional tiers, but the extent of reform varies by a factor of twenty. The variable *Reform* is calculated as the cumulative absolute change in the RAI country score going back in time, so that values are maintained or increase as one moves back from the present.

Analysis of Disagreement

To what extent do these potential biases explain disagreement between the RAI and prior measures of decentralization? Our strategy is to extract residuals

[7] The logic is that retrospective judgments may be more unreliable (Steenbergen and Marks 2007). We calculate this variable for each measure.

Table 2.4. Explaining absolute disagreement

Source of disagreement	Arzaghi and Henderson	Brancati	Lijphart	Treisman	Woldendorp et al.
Limited coverage	0.022	0.015	0.002	0.049**	0.010
	(0.016)	(0.006)	(0.012)	(0.009)	(0.010)
Distance in time	0.009**	0.015**	0.005**	0.015**	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Tiers	0.147**	0.145**	0.236**	0.121**	0.191**
	(0.036)	(0.030)	(0.038)	(0.026)	(0.031)
Differentiation	0.002	−0.008**	0.004	0.001	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Reform	−0.021**	0.023**	−0.005	0.041**	−0.016*
	(0.004)	(0.006)	(0.007)	(0.006)	(0.006)
Rho	0.875	0.928	0.965	0.899	0.968
R^2	0.16	0.34	0.13	0.58	0.16
Wald chi^2	237	391	133	742	168
N years	1030	847	1178	606	1137
N countries	29	58	27	63	32

Note: * $p < 0.05$, ** $p < 0.01$ (two–tailed).

OLS regressions with panel corrected standard errors in brackets. The constant is dropped. The dependent variables are the *absolute* standardized residuals resulting from an OLS regression of the RAI on one of the decentralization indices.

for disagreement between the RAI and each of the five decentralization measures by regressing the RAI on the country/year scores generated by each measure.[8] We then regress the standardized residuals onto the variables discussed above using ordinary least squares regression with panel corrected standard errors.[9]

Error comes in two forms (Marks et al. 2007). Absolute residuals capture the sheer distance between scores. This gives us a sense of how far a measure strays from other measures regardless of the direction of the difference. Raw residuals come with signs that tell one the direction of difference between scores, i.e. whether a score is in the direction of more or less decentralization. Table 2.4 presents models explaining absolute residuals and Table 2.5 does the same for raw residuals.

Distance in time, *Tiers*, and *Reform* are consistently positive causes of difference between the RAI and prior measures. The further back in time one estimates decentralization, the greater the number of levels of regional governance, and the greater the extent of jurisdictional reform over time, the larger the discrepancy between the RAI and the alternative measures. *Tiers* has the most marked effect. The absolute difference in scoring between Lijphart

[8] Table 2.A.4 reports Pearson correlations between the residuals.

[9] Table 2.A.5 displays descriptive statistics for the independent variables. See Achen (2000); Plümper et al. (2005); and Beck and Katz (2011), for a discussion of the conditions under which panel corrected standard errors without a lagged dependent variable and fixed effects are appropriate.

Table 2.5. Explaining directional disagreement

Source of disagreement	Arzaghi and Henderson	Brancati	Lijphart	Treisman	Woldendorp et al.
Limited coverage	0.066**	−0.018**	−0.005	−0.051**	−0.003
	(0.020)	(0.006)	(0.009)	(0.011)	(0.007)
Distance in time	−0.013**	−0.012**	−0.001	−0.015**	0.000
	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
Tiers	0.162**	0.281**	0.250**	0.333**	0.202**
	(0.048)	(0.033)	(0.027)	(0.036)	(0.022)
Differentiation	0.025**	−0.000	0.002	0.005	−0.001
	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)
Reform	−0.006	−0.049**	−0.066**	−0.028**	−0.069**
	(0.006)	(0.008)	(0.004)	(0.010)	(0.004)
Rho	0.907	0.940	0.966	0.929	0.971
R^2	0.08	0.34	0.50	0.206	0.56
Wald chi^2	68	312	550	173	803
N years	1030	847	1178	606	1137
N countries	29	58	27	63	32

Note: * $p < 0.05$, ** $p < 0.01$ (two–tailed).

OLS regressions with panel corrected standard errors in brackets. The constant is dropped. The dependent variables are the *raw* standardized residuals resulting from an OLS regression of the RAI on one of the decentralization indices. A positive sign indicates that the estimate of the RAI is higher than the estimate of the alternative measure.

and the RAI for a country, such as Finland, Italy, or Portugal, which have two tiers of regional governance instead of one, would on average be just under a quarter (23.6 percent) of a standard deviation in the RAI score, which is equivalent to a score difference of 2.2. Estimates for the effect of *Distance in time* and *Reform* are greatest for the RAI and Treisman. A country/year scored twenty years in the past would, on average, generate a difference in scoring of around one-third of a standard deviation—or around 3.2 on the RAI scale.

An examination of the results for directional disagreement provides some meat on these bones. The most notable result is that the RAI detects more decentralization than alternative measures in the presence of multiple levels of regional government. This is precisely what one would expect given that the RAI estimates each level prior to aggregating them to the country level, whereas the other measures do not distinguish multiple levels in estimating decentralization. The substantive effect is quite marked. The RAI assesses between 16 percent and a third of a standard deviation more decentralization than the remaining measures in a country that has a second tier of regional government compared to just one.

Reform also has the anticipated effect. In general, jurisdictional reform has increased the level of decentralization over the past several decades. The RAI estimates lower levels of decentralization in past years where jurisdictional reform—and hence the increase in decentralization—has been large. The difference is not significant compared to the Azarghi and Henderson measure, which

picks up change at five-year instead of annual intervals. However, it reduces the assessment of decentralization by around two-thirds of a standard deviation compared to Lijphart and Woldendorp et al. for the period prior to decentralization reforms in, say, Greece, which empowered its deconcentrated *nomoi* to self-governing units in the 1990s, when a second intermediate tier of *periphereies* was introduced. These reforms increased Greece's country score by ten points.

Cases of Disagreement

Outlying cases can be particularly revealing. So let us take a closer look at cases where the residual is more than two standard deviations above or below the estimate for five or more consecutive years. Table 2.6 lists twenty-one such cases in ten countries. Residuals with a positive sign are those where the RAI estimate is higher than the alternative measure. Disagreement is often greatest when a country has multiple tiers of regional government; when there is considerable variation in decentralization over time; or when there

Table 2.6. Cases of disagreement

Country	Years	(Range of) z–scores	Measurement
Belgium	1989–2000	+2.20/+2.74	Brancati
	1980–1994	+2.24/+3.40	Lijphart
	1990–1994	+2.01	Treisman
	1980–1998	+2.06/+3.13	Woldendorp et al.
Chile	1960–1974	–2.01	Arzaghi and Henderson
	1995–1999	–3.12	Arzaghi and Henderson
Finland	1950–1992	–2.06	Woldendorp et al.
France	1982–1996	+2.24/+2.40	Lijphart
	1990–1999	+2.25	Treisman
Germany	1977–1989	+2.05	Arzaghi and Henderson
	1985–2000	+2.52/+2.84	Brancati
	1990–1999	+2.43/+2.47	Treisman
	1977–1989	+2.01	Woldendorp et al.
Italy	1989–1996	+2.13/+2.43	Lijphart
Serbia and Montenegro	1992–2000	+2.54/+2.83	Brancati
	1992–1999	+2.15/+2.48	Treisman
Spain	1978–1999	+2.14/+3.15	Arzaghi and Henderson
	1983–1996	+2.39/+2.86	Lijphart
	1983–1998	+2.12/+2.70	Woldendorp et al.
Trinidad and Tobago	1985–1995	–2.02	Brancati
Venezuela	1950–1960	–2.70/–2.46	Lijphart

Note: A case of disagreement is defined as two standard deviations below or above the estimate for a time period of five years. A positive sign indicates that the estimate of the RAI is higher than the estimate of the alternative measure.

is differentiation. In some cases, different scores reflect more fundamental differences in conceptualization and operationalization.

Belgium and Germany stockpile the largest number of disagreements. The RAI scores diverge with four of the five alternative measures and in each case the RAI score is higher. For Belgium, the single most important factor is the cadence of reform—five major reforms between 1970 and 2005. Static measures such as Treisman or Woldendorp et al. are poorly equipped to capture this. Lijphart's measure is not entirely static since he increases Belgium's score from 3.1 to 5 in 1993 following federalization. However, Lijphart's measure does not pick up the regional empowerment that took place in the 1970s. The divergence with Treisman for 1990 to 1994 reflects a scoring disagreement: the Belgian senate does not meet Treisman's criterion for a regional chamber, while it does according to the RAI.

Brancati's measure, which is the only one to provide annual readings between 1985 and 2000, registers no change in Belgium, whereas the RAI spikes up in 1989 when Belgian regions and communities obtain broader policy competences, taxation powers, and shared rule. This alerts us to a difference in conceptualization. Brancati's measure emphasizes electoral and policy autonomy, but the central foci of the 1989 and 1993 reforms were tax autonomy, executive federalism, and a reform of the senate.

Disagreement between the RAI and the alternative measures in estimating decentralization in Germany appears to result from conceptual differences between the RAI and these measures. The RAI evaluates multiple tiers, and it pays close attention to shared rule; Germany has both multiple levels of regional governance and high levels of shared rule. The RAI picks up the authority exercised by regional governments within *Länder* (including *Regierungsbezirke* and *Kreise*) and it considers several dimensions of shared rule, including intergovernmental meetings between *Länder* and the federal government.

Disagreement with Treisman and Brancati also reflects coding judgments. Treisman's score of 1.5 out of a possible 3.0 for Germany is based on a restrictive interpretation of *Länder* authority: the absence of constitutionally entrenched exclusive powers, and the absence of an absolute veto by the *Bundesrat* on legislation (though it can raise the hurdle). The RAI, by contrast, considers concurrent powers and the role of *Länder* in implementing national framework legislation (Swenden 2006; Watts 1999*a*). Brancati scores Germany 3.0 out of a possible 5.0 because she estimates that constitutional amendments do not require *Länder* approval. The RAI registers that *Länder* have a veto on constitutional reform by virtue of their representation in the *Bundesrat*.

The Lijphart index disagrees with the RAI for three more countries: Venezuela, France, and Italy. Lijphart scores Venezuela significantly higher than

the RAI for the 1950s. Venezuela receives a score of 4 out of 5 on the Lijphart index, which is consistent with the 1947 constitution for a centralized federation, but this constitution was never put into effect due to a military coup in 1948, and the new constitution of 1952 replaced elected by appointed officials at all levels (1948–57).[10] The RAI also scores Italy after its 1989 reform and France after the Defferre reform of 1982 as having more decentralization than in Lijphart's measure.

Chile and Finland have higher scores in Arzaghi and Henderson and Woldendorp et al., respectively, chiefly because these authors include local government in their measure. The RAI estimates Chilean *provincias* and *regiones*, which are primarily deconcentrated, whereas Arzaghi and Henderson code municipal authorities as not being subject to central veto. The Woldendorp et al. measure scores Finland higher because it captures Finland's relatively authoritative municipal authorities while the RAI does not. From 1993, when Finland creates self-governing regional governments, which are picked up by the RAI, the two indices fall in line.

Three indices estimate Spain to have considerably less decentralization than the RAI. In contrast to Lijphart and Woldendorp et al., the RAI encompasses scores for *provincias* as well as *comunidades autónomas*. Arzaghi and Henderson consider both levels of governance, but their score is subdued because they focus on primary education, infrastructure, and policing—areas in which the central government retained substantial authority.

The RAI and Brancati differ on Serbia and Montenegro and Trinidad and Tobago on definitional grounds. Whereas Brancati scores Serbia, the RAI scores the federation and, from 2003, the "state union" of Serbia and Montenegro. Trinidad and Tobago consists of two main islands but only Tobago has an intermediate tier of government. Brancati's score for Tobago is the same as for the country as a whole, while the estimate of the RAI is lower because the score for Tobago is weighted by its population size.[11]

Two remaining cases of disagreement with Treisman are France and Serbia and Montenegro. Treisman gives France a score of zero because his coding registers only constitutional provisions, while the authority exercised by *départements* and *régions* is laid down in special legislation. Serbia and Montenegro has a score of 1 on a scale from zero to 3, which is surprisingly low for a (con) federation. Again, Treisman's emphasis on constitutional criteria explains this. Serbia and Montenegro's upper chamber is not coded as regional, probably because it was not directly elected but made up of twenty deputies from each member republic. Instead, the RAI registers extensive shared rule through the

[10] For greater detail, see the country profile of Venezuela.
[11] Tobago's population is 60,000 and that of the country as a whole is 1.3 million (2011 figures).

upper chamber, giving Serbia and Montenegro one of the highest scores in the RAI dataset.

Fiscal Indicators

Fiscal indicators are widely employed in studies of decentralization (see e.g. Blöchliger 2015; Blöchliger and King 2006; Braun 2000; Castles 1999; Harbers 2010; Oates 1972; Stegarescu 2005*a*; Willis et al. 1999). The principal sources are Government Finance Statistics (GFS) produced by the International Monetary Fund (IMF) and Historical National Accounts and Revenue Statistics produced by the OECD.[12] Authors interested in the effects of decentralization on outcomes such as economic growth, corruption, or redistribution, have used revenue and expenditure indices in combination (Akai and Sakata 2002; Enikolopov and Zhuravskaya 2007; Jin and Zou 2002). Some authors have sought to increase the validity of specific fiscal indicators (Ebel and Yilmaz 2002; Stegarescu 2005*b*).

Despite these efforts, two basic caveats remain when using fiscal indicators to tap regional authority (Blöchliger 2015; Rodden 2004; Schakel 2008; Sorens 2011). The first is that the extent of subnational expenditure or revenue does not indicate the autonomy of a subnational government from central control in spending money. *Departamentos* in Uruguay, for example, spend more than twice as much as a proportion of total government expenditure than those in Bolivia (15.4 percent versus 7.2 percent), but have less authority over taxes (Daughters and Harper 2007: 224). Subnational governments in South Korea were conduits for 34.4 percent of total government expenditure in 1978 (the latest year reported in World Bank data) at a time when the country was highly centralized under military rule. In the same year, popularly elected Malaysian subnational governments with diverse policy making powers were responsible for 17.2 percent of total government expenditure, and subnational governments in Indonesia, which were more authoritative than those in South Korea, spent just 13.4 percent of total government expenditure.

The amount a government spends does not tell us whether spending is financed by conditional or unconditional grants, whether the central government determines how the money should be spent, or whether it sets the framework legislation within which subnational governments implement (Blöchliger 2015; Akai and Sakata 2002; Breuss and Eller 2004; Ebel and Yilmaz 2002; Fisman and Gatti 2002; Martinez-Vazquez and McNab 1997; Panizza 1999). Figure 2.1 shows that subnational governments in

[12] We use the World Bank (2006) Fiscal Indicator dataset derived from the GFS (IMF) because it has the greatest overlap with the RAI: fifty-six countries with yearly scores for 1972–2000.

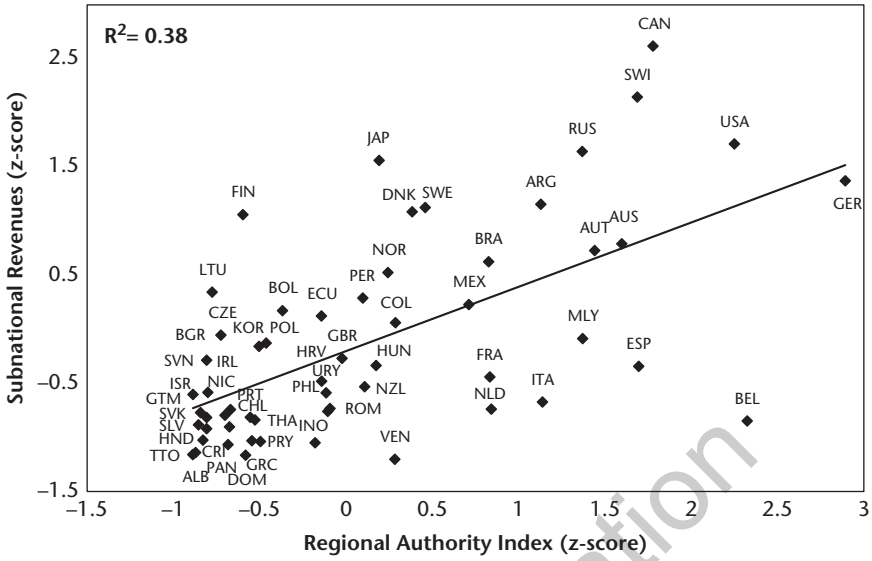

Figure 2.1. Subnational expenditure and regional authority

Note: Subnational expenditure as a percentage of total government expenditures. World Bank (2006) Fiscal Indicator dataset. Standardized scores for country means for 1972–2001.

Scandinavian countries have the same (or higher) shares of total government expenditures than their peers in federal countries. However, subnational governments in Scandinavian countries have less decision making authority over policies, less taxation power, and they do not enjoy power sharing. The national government usually determines the policies that are implemented by local and regional governments.

Subnational revenue consists of tax and non-tax revenue (e.g. fees, receipts, and levies), intergovernmental transfers, and other grants. A recurring debate concerns the classification of sources of subnational income. For example, revenues from shared taxes are assigned to subnational governments in the GFS database even when subnational governments have no autonomy over the revenue base or rate. This has led scholars to develop revenue indicators for "own" subnational revenue and subnational tax autonomy. "Own" subnational revenue is the ratio of revenue, exclusive of received intergovernmental transfers, to total subnational revenue. Subnational tax autonomy consists of taxes that can be determined by subnational government and which are subject to subcentral legislative and administrative powers (Ebel and Yilmaz 2002; Stegarescu 2005*b*).

The fiscal envelope of a subnational government does not capture the authority of a government to regulate behavior. This is the distinction between "regulatory policies and policies involving the direct expenditure of public funds" (Majone 1994). Some policies, including redistributive policies, have a

direct bearing on the public budget, whereas regulatory policies, including civil and criminal law, may have considerable impact on society by virtue of the rules they impose. While the cost of expenditure programs is borne by the public budget, the cost of most regulatory policies is borne by citizens and firms (Majone 1994). To the extent that regions have control over regulatory policies, expenditure fiscal indicators reveal little about decentralization. Policy authority is captured by the RAI separately from fiscal authority.

Scholars have also produced measures for central grants to subnational governments (Akai and Sakata 2002; Oates 1972; Stegarescu 2005*b*). Vertical imbalance is the degree to which subnational governments rely on central government revenues to support their expenditures, and is measured by intergovernmental transfers as a share of subnational expenditures. This has been criticized because it does not identify whether a grant comes with a centrally imposed mandate (Shah 2007). This is a valid concern, but we do not have reliable data that distinguish between conditional and unconditional grants (Rodden 2004).

A second caveat is that fiscal indicators provide one score for all levels of subnational government. Fiscal decentralization indices do not distinguish between local and regional tiers and do not take differentiated governance into account.[13] Further, the existence of regional governments with special powers can shape the level of decentralization in a society even if their powers are not generalized across an entire tier of subnational government. For example, the Basque *foru iurralde* (historic territories) and Navarre in Spain collect income, corporate, inheritance, and wealth taxes and can set the rate and base for these taxes autonomously, whereas in the rest of Spain the bulk of taxes are paid to the center and set amounts are transferred back to the regions (Swenden 2006). The five special *regioni* and the *provinces* of Bolzano-Bozen and Trento in Italy receive a share of taxes collected in their jurisdictions whereby the central government sets the base but the rate is negotiated bilaterally between the region and central government. In contrast, the tax autonomy of ordinary *regioni* is limited: they can set the rate within centrally determined limits for minor taxes (the vehicle tax, an annual surtax, a special tax on diesel cars, and health taxes). Failing to capture this variation can over- or underestimate fiscal autonomy.

Figure 2.2 plots subnational revenue as a percentage of total government revenue against the RAI country score. The correlation is statistically significant, but a closer look reveals that Sweden and Denmark are ranked on a par with Argentina and have higher scores than Australia, Austria, Brazil, and Mexico. Counties in Denmark and Sweden may set the rate of income tax within central government parameters but it would be wrong to conclude that

[13] As before, we exclude countries without a regional tier.

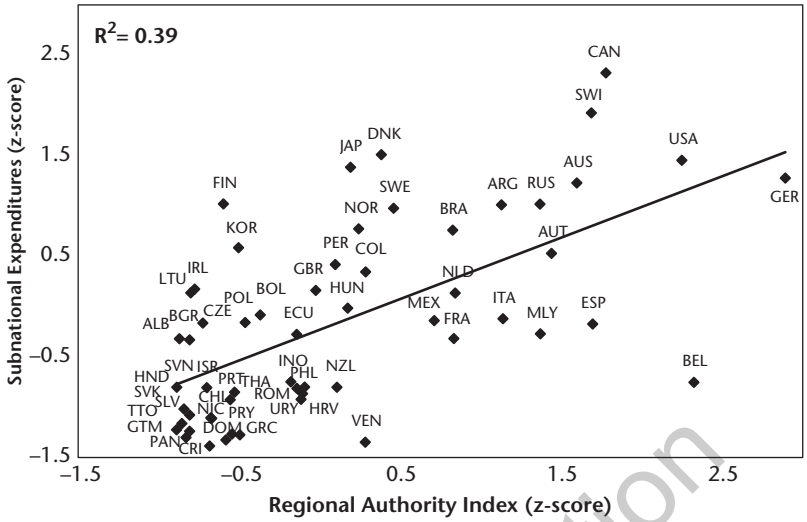

Figure 2.2. Subnational revenue and regional authority

Note: Subnational revenue as a percentage of total government revenue plotted against country scores on the RAI. World Bank (2006) Fiscal Indicator dataset. Standardized scores for averages for 1972–2001.

the subnational tiers in Sweden and Denmark enjoy the same autonomy as their peers in these federal countries.

One way to gain more insight into central involvement in subnational revenue and expenditure is by looking at the share of intergovernmental grants (Akai and Sakata 2002; Blöchliger 2015; Breuss and Eller 2004; Oates 1972; Stegarescu 2005*b*). A common measure is vertical imbalance, which is operationalized by intergovernmental transfers as a share of subnational expenditures. However, this indicator is also limited. Aside from data availability regarding unconditional and conditional grants (Rodden 2004), there is the problem that intergovernmental grants do not seem to differentiate between federal and non-federal countries.

Figure 2.3 displays vertical imbalance against RAI country scores. One would expect a negative relationship between vertical imbalance and RAI scores since high percentages of central government grants relative to total subnational revenue should be associated with low scores on the RAI. As one can observe in Figure 2.3 the vertical imbalance in decentralized federal countries such as Canada, Switzerland, and the US is comparable to that in centralized unitary countries such as Bulgaria, Lithuania, and Slovakia.[14]

[14] An analysis of variance (ANOVA) shows that the extent of vertical imbalance does not vary significantly between unitary and federal countries (F: 1.99; df = 52, p = 0.147).

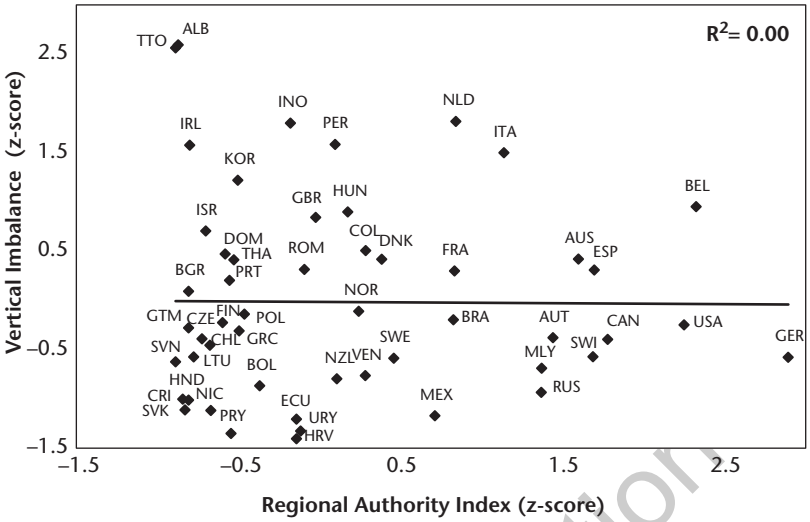

Figure 2.3. Vertical imbalance and regional authority

Note: Intergovernmental transfers as a share of subnational expenditures plotted against country scores on the RAI. World Bank (2006) Fiscal Indicator dataset. Standardized scores for averages for 1972–2001.

Fiscal measures conflate meaningful decentralization with change in public spending (Stegarescu 2005*b*). Fiscal decentralization may differ between two countries even in the case of an identical allocation of policies and functions across tiers of government (Oates 1972; Panizza 1999). A country that spends relatively more on policies that are centralized for scale efficiency reasons (such as defense) will also be more fiscally centralized.[15] A similar argument applies to the question whether welfare state policies are provided by the government or by the private sector. For example, in the Scandinavian countries, a large proportion of government expenditure goes to welfare state policies and these are often provided by subnational governments. In market-liberal Anglo-Saxon countries, many welfare state functions are privatized. So a difference in political economy leads to higher expenditure (and revenue) in Scandinavian countries than in Anglo-Saxon countries, whereas the allocation of functions among levels of government may be identical.

[15] The World Bank (2006) notes that particular expenditure categories can distort measures of decentralization: "For instance, the United States, despite being a much larger country, has a lower sub-national share of expenditures than Switzerland. However, when defense and interest expenses are excluded from the subnational-to-total ratio, the United States has a higher subnational share of expenditures than Switzerland." Some scholars therefore use fiscal indicators that exclude defense, e.g. Breuss and Eller (2004) and Panizza (1999).

The RAI avoids conflating these distinct political processes by measuring autonomy rather than expenditure.

Conclusion

The extent to which states are decentralized has been a topic of enduring interest in political science. Measures of decentralization are used in explaining a wide range of political outcomes concerning public policy, the quality of governance, and economic performance. Scholars have employed institutional indicators of decentralization or used fiscal indicators based on World Bank data on subnational expenditures and revenues.

This chapter crossvalidates the RAI with five commonly used institutional measures, and finds much agreement. Notwithstanding their marked differences in conceptualization, operationalization, and coverage, a single underlying factor accounts for more than 79 percent of the variance. Decentralization appears to have a core meaning that can be tapped by measures using very different indicators. But we also detect systematic sources of disagreement and divergent interpretation of evidence.

The most consistent sources of systematic disagreement arise from the fact that the RAI is better equipped to account for multilevel regional governance and for regionalization over time because it estimates the authority of subnational governments at multiple levels and because it produces annual observations over six decades. Beyond the systematic differences among these measures, there are numerous differences in the interpretation of particular cases. In this chapter we assess these on a case by case basis drawing on documentary evidence.

The associations between the RAI and World Bank data on subnational expenditures and revenues are relatively weak at 0.60 and 0.59, respectively. In their current form, fiscal indicators are not good at capturing whether regional governments decide autonomously over revenues and expenditures, nor do they encompass authority over regulation that does not involve much money.

The comparisons among the available measures in this chapter suggest that there is non-negligible consensus among experts in estimating decentralization at the country level. However, the purpose of the RAI is to measure authority at the regional level, a more difficult and perhaps more hazardous undertaking. Our concern in the next chapter is with accuracy rather than consensus, and this leads us to engage the substantive content of the RAI, probing the theoretical, conceptual, and operational decisions that underpin it.

Appendix

Table 2.A.1. Country coverage across measures of decentralization

	A-H	BRA	LIJP	TRE	WKB	Total
Albania				X		1
Argentina	X	X		X		3
Australia	X		X	X	X	4
Austria		X	X	X	X	4
Belgium		X	X	X	X	4
Bolivia		X		X		2
Bosnia and Herzegovina		X		X		2
Brazil	X	X		X		3
Bulgaria		X		X	X	3
Canada	X	X	X	X	X	5
Chile	X	X		X		3
Colombia	X	X	X	X		4
Costa Rica		X	X	X		3
Croatia		X		X		2
Cuba				X		1
Czech Republic		X		X	X	3
Denmark		X	X	X	X	4
Dominican Republic		X		X		2
Ecuador	X	X		X		3
El Salvador		X		X		2
Finland		X	X	X	X	4
France	X	X	X	X	X	5
Germany	X	X	X	X	X	5
Greece	X	X	X	X	X	5
Guatemala		X		X		2
Haiti				X		1
Honduras		X		X		2
Hungary	X	X		X	X	4
Indonesia	X	X		X		3
Ireland		X	X	X	X	4
Israel		X	X	X	X	4
Italy	X	X	X	X	X	5
Japan	X	X	X	X	X	5
Lithuania		X		X	X	3
Malaysia	X	X		X		3
Mexico	X	X		X		3
Netherlands	X	X	X	X	X	5
New Zealand		X	X	X	X	4
Nicaragua		X		X		2
Norway		X	X	X	X	4
Panama		X		X		2
Paraguay		X		X		2
Peru	X			X		2
Philippines	X	X		X		3
Poland	X	X		X	X	4
Portugal		X	X	X	X	4
Romania	X	X		X	X	4
Russia	X	X		X		3
Serbia and Montenegro		X		X		2
Slovakia		X		X	X	3

(*continued*)

Table 2.A.1. Continued

	A-H	BRA	LIJP	TRE	WKB	Total
Slovenia		X		X		2
South Korea	X	X		X		3
Spain	X	X	X	X	X	5
Sweden		X	X	X	X	4
Switzerland		X	X	X	X	4
Thailand	X	X		X		3
Trinidad and Tobago		X	X	X		3
Turkey	X	X		X	X	4
United Kingdom	X	X	X	X	X	5
United States	X	X	X	X	X	5
Uruguay		X		X		2
Venezuela	X	X	X	X		4
Total	29	57	26	62	30	204
Coverage	47%	92%	42%	100%	48%	66%

Note: Nineteen countries without a region or regional tier are excluded.
A-H = Arzaghi and Henderson; BRA = Brancati; LIJP = Lijphart; TRE = Treisman; WKB = Woldendorp, Keman, and Budge.

Table 2.A.2. Descriptive statistics

Measurement	Mean	St.dev.	Min	Max
Regional Authority Index	9.62	9.35	0.00	36.95
Arzaghi and Henderson	1.93	1.21	0.00	4.00
Brancati	2.09	1.13	0.00	5.00
Lijphart	2.56	1.54	1.00	5.00
Treisman	0.51	0.86	0.00	3.00
Woldendorp, Keman, and Budge	3.40	1.92	0.00	7.00
World Bank subnational expenditure	23.31	15.21	1.45	59.18
World Bank subnational revenue	17.27	13.73	0.13	54.60
World Bank vertical imbalance	36.96	22.82	0.29	96.60

Table 2.A.3. Pairwise Pearson correlations

	1	2	3	4	5	6	7	8	9
1	1.00								
2	0.84**	1.00							
3	0.73**	0.67**	1.00						
4	0.78**	0.78**	0.71**	1.00					
5	0.77**	0.63**	0.71**	0.69**	1.00				
6	0.77**	0.86**	0.73**	0.80**	0.67**	1.00			
7	0.60**	0.71**	0.50**	0.56**	0.49**	0.79**	1.00		
8	0.59**	0.67**	0.56**	0.62**	0.50**	0.82**	0.94**	1.00	
9	–0.02	–0.25**	–0.09*	–0.27**	–0.09	–0.42**	0.05	–0.27**	1.00

Note: * $p < 0.05$; ** $p < 0.01$.
1 = RAI; 2 = Arzaghi and Henderson; 3 = Brancati; 4 = Lijphart; 5 = Treisman; 6 = Woldendorp, Keman, and Budge; 7 = World Bank subnational expenditure; 8 = World Bank subnational revenue; 9 = World bank vertical imbalance.

Table 2.A.4. Pairwise Pearson correlations between the residuals of the alternative measurements

		1	2	3	4	5
1	Arzaghi and Henderson	1.00				
2	Brancati	0.47	1.00			
3	Lijphart	0.56	0.71	1.00		
4	Treisman	0.34	0.62	0.56	1.00	
5	Woldendorp, Keman, and Budge	0.70	0.73	0.70	0.59	1.00

Note: all Pearson correlations are statistically significant at the $p < 0.0001$ level. The residuals are obtained by regressing the decentralization index on the RAI.

Table 2.A.5. Descriptive statistics of the independent variables

Variable	Mean	St.dev.	Min	Max
Limited coverage	3.17	1.32	0.00	5.00
Distance in time	35.00	14.93	11.00	61.00
Tiers	1.14	0.50	0.00	2.86
Differentiation	4.80	7.04	0.00	24.00
Reform (A-H)	6.04	8.47	0.00	58.26
Reform (BRA)	1.67	2.55	0.00	13.92
Reform (LIJP)	3.59	4.69	0.00	29.26
Reform (TRE)	1.02	1.84	0.00	9.00
Reform (WKB)	3.62	4.91	0.00	30.20

Note: A-H = Arzaghi and Henderson; BRA = Brancati; LIJP = Lijphart; TRE = Treisman; WKB = Woldendorp, Keman, and Budge.