

# **Taxation Data as Indicators of State-Society Relations: Possibilities and Pitfalls in Cross-National Research\***

*Evan S. Lieberman*

Cross-national research on taxation is a growth industry in political science. This article discusses key conceptual and measurement issues raised by such studies. First, it highlights the ways in which taxation has been studied as a rich and varied concept, including as a component of the state-building process, as a collective action problem, and/or as a problem of distributive justice. Second, the article identifies the central tradeoffs associated with the construction of taxation indicators used to measure such ideas. It discusses considerations such as which forms of revenue should be included and which should not, whether and how to standardize taxation measures, and how to fine-tune measures through a clear specification of units, universes, and measurement calibration. These choices have important implications for the “scoring” of countries, and for making valid inferences about the relationship between states and societies.

## **I. Introduction**

Political scientists are increasingly carrying out cross-national research on taxation in order to explore many of the central dilemmas and problems of modern political life. Since the late 1980s, taxation has figured prominently in several important cross-national studies published in political science journals and books (Levi 1988; Peters 1991; Steinmo 1993, 1998; Garrett 1998; Chaudhry 1997; Cheibub 1998; Fauvelle-Aymar 1999), creating an emerging sub-field of the comparative political-economy of taxation. Theoretically and empirically, taxation provides an attractive focus for research, as the real world challenges associated with tax collection are intimately linked with many of the central analytic concerns of political scientists, including questions about state building, collective action, and distributive justice. By making links be-

---

Evan S. Lieberman is a Robert Wood Johnson Policy Scholar at Yale University and will begin an appointment as assistant professor in the Department of Politics at Princeton University in September 2002. His interests include comparative politics, research methods, and economic and social policy.

tween these constructs and the process of tax collection, scholars have found that it is possible to take advantage of a wealth of readily available, highly-standardized taxation datasets that are amenable to comparative analysis, both across countries and over time.

While the availability of such taxation data does facilitate the cross-national study of these larger problems, there is no clear consensus about how various constructs should be measured, or how to interpret various indicators (Table 1). For example, Cheibub's essay (1998) relies exclusively on a single indicator—total taxes as share of GDP, whereas Peters' book uses more than 20 indicators to identify cross-national variation in tax structures. Steinmo (1993) measures various forms of tax collection as a proportion of GDP *and* of total revenues. By contrast, Chaudhry's (1997) and Levi's (1988) respective studies largely employ indicators that report tax collections in local currencies, and not in relation to some other measure. How are we to interpret these differ-

**Table 1**  
**Examples of Tax Revenue Indicators**  
**in the Field of Comparative Politics**

Author	Tax Indicators Employed in Study			
	Numerator	Denominator	Unit	Universe
Chaudhry (1997)	Tax on wages Corporate profit tax Zakat Total direct Indirect taxes Customs Other Stamp duties Total domestic, no customs Total domestic Direct taxes Indirect taxes	none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) none (local currency) GDP GDP	Central state Central state Central state Central state Central state Central state Central state Central state Central state Central state Central state Central state Central state	Middle East; Late, Late Developers
Cheibub (1998)	Total tax	GDP	All levels	All countries
Fauvelle-Aymar (1999)	Total tax Current revenue	GDP GDP	Central state Central state	All countries
Garrett (1998)	Total tax Personal income tax Consumption taxes Corporate income tax Soc sec contributions	GDP GDP GDP GDP GDP	All levels All levels All levels All levels All levels	Advanced countries
Levi (1988)	Total tax Direct tax	none (local currency) Total revenue	Varied Central state	Varied
Peters (1991)	Personal income Corporate income Employees' social security Employers' social security Payroll Property Wealth Other Total tax Social security Income	GDP, Total revenue GDP, Total revenue GDP GDP GDP GDP GDP GDP GDP GDP Total revenue Total revenue	All levels All levels All levels All levels All levels Varied All levels All levels All levels All levels All levels Local	Advanced countries
Steinmo (1993)	Total tax Social security Income and profits Consumption Property Direct	GDP, Total revenue GDP, Total revenue GDP, Total revenue GDP, Total revenue GDP, Total revenue GDP, Total revenue	Varied Varied Varied Varied Varied Varied	Advanced countries
Steinmo (1998)	Total tax	GDP	All levels	Advanced countries

ences? What are the consequences of using one indicator versus another? Insufficient attention has been paid to conceptual and measurement issues in these studies,<sup>1</sup> making it difficult for scholars to assess, replicate, and/or extend such work. Clearly, if there is bias in the indicators, then causal assessments of the relationships *between* concepts will necessarily be biased as well. Too often, government revenue statistics are presented as if they were transparent representations of social reality, without explication of how they should be interpreted, or of the potential for measurement error.

This article discusses the problem of measurement validity (Adcock and Collier 2001) as it applies to cross-national studies of taxation. The primary goal is to achieve greater conceptual clarity by unpacking taxation indicators, and by reflecting upon the political implications of their component parts. It analyzes the degree to which various tax indicators can serve as valid and reliable measures of different sets of constructs and for different universes of cases. In the same manner that survey responses can provide useful-but-imperfect approximations of individual attitudes and traits, taxation data can be used as a measurement instrument for societal-level analyses. Measurement strategies have advantages and disadvantages, and scholars need to make choices with respect to how well a particular interpretation of tax collections lines up with the ideas or concepts being studied. The article provides a framework for making such choices by identifying possible conceptual interpretations of tax collection and by specifying the various assumptions, tradeoffs, and considerations associated with such theoretical linkages.

## II. Conceptualizing Taxation

Perhaps because taxation is generally more closely associated with economics<sup>2</sup> or because taxation has become so central to everyday life, it is often treated as an unproblematic concept. Levels of tax collections are of intrinsic interest simply because they are a key source of government revenue that provides funding for welfare, defense, and other government programs around the world, and because of their influence on markets. Notwithstanding, scholars of comparative politics have tended to focus on taxation because they believe it reveals important dimensions of political life that cannot be directly observed. They have described the imperative to collect revenue as a political “problem” that gets resolved in varied ways across space and time.

Looming in the background of most taxation studies is some general agreement about what defines taxation in theory and in practice. Taxation is a form of government revenue that differs from other forms of finance, including debt, entrepreneurial (parastatal) income, and user fees, in terms of obligations and administrative requirements. Taxes are “unrequited compulsory payments collected primarily by the central government” (World Bank 1988: 79). They are levied on a particular base and paid to the government to provide certain public goods or services or to redistribute income or purchasing power within society, but without provision or promise of any *specific* good or service in return for payment. As Lorenz Von Stein explains in his classic work on taxation:

Taxes are conceptually entirely different from all other public revenue... Taxes can be said to represent the nation's entire civic sense on the economic plane.... In administering public property, the State is an independent economic agent with its own capital; fees and regalia represent a payment to the State in return for services rendered to individuals for the satisfaction of their individual needs. Taxation, by contrast, represents a field in national economic life where, by virtue of the State's constitution and administration, part of the individuals' economic income is withdrawn from them and becomes the community's economic income... (Von Stein [1885] 1964: 28)

Implicitly or explicitly, many political scientists have taken up the project of "fiscal sociology," which employs taxation as a useful lens onto the relationship between states and societies. Seminal essays by Joseph Schumpeter and Rudolph Goldscheid provided strong scholarly foundations for studying taxation as a central institution in social and political life and helped spawn this still nascent field. In his essay on the "tax state," Schumpeter explained, "The spirit of a people, its cultural level, its social structure, the deeds its policy may prepare—all this and more is written in its fiscal history.... The public finances are one of the best starting points for an investigation of society, especially though not exclusively of its political life" (Schumpeter [1918] 1954: 7). Reflecting on more recent contributions, Campbell explains that fiscal sociology focuses "explicitly on the complex social interactions and institutional and historical contexts that link state and society in ways that shape fiscal policies and their effects" (Campbell 1993: 164).

Political scientists have begun to stake some ground in this line of research by offering interpretations of tax collections and novel theories that account for cross-national and over-time variation. Principally, three key concepts have been central to the study of the political economy of taxation: state capacity, collective action, and distributive justice. In each case, scholars have made the case that the specific task of tax collection provides a lens for understanding more general processes of political life and the relationship between state and society.

### *A. Emergence, Size, and Capacity of the State*

The most widespread interest in taxation within the field of comparative politics has been in terms of the close relationship between the capacity to collect revenues and the construction of the modern national state. The development of state power, or the state's authority over society and the market economy, is usefully examined by highlighting its ability to get citizens to do something that they would rather not do—namely, pay taxes. Tax collection is ultimately the product of policy making, the monitoring of economic activity, the administration of complex laws, and judicial and punitive enforcement. For scholars, varied levels of tax revenues reflect variations in these state processes.

Several seminal studies have used taxation to study the state in this manner. Douglass North goes so far as to *define* the state in terms of taxation powers: "... an organization with a comparative advantage in violence, extending over

a geographic area whose boundaries are determined by its power to tax constituents” (North 1981: 21). Several contributions in Charles Tilly’s (1975) classic edited volume of the emergence of national states in Western Europe also describe taxation as intricately linked to the emergence of modern state forms, a theme that Tilly himself builds upon in a later work (Tilly 1992).

More recent contributions have followed this line of theorizing by using taxation as a way of measuring the size and scope of the state around the world and over vast expanses of time. For example, Chaudhry’s analysis of the development of the Saudi and Yemeni states depicts tax collection as an important aspect of state building, specifically within the context of “late late” development (Chaudhry 1997: 25-6, 32-4). She points out that the regulatory and information-gathering dimensions of the state get formed in the process of extending the taxation apparatus, implying that actual tax collections reflect more broadly on the development of these other bureaucracies. Similarly, Cheibub uses taxation as a way of measuring the ability of governments to impose unpopular policies, a construct he calls, “the extractive capacity of governments” (Cheibub 1998: 350). Following Levi (1988), he highlights the need of governments to monitor and to enforce contracts as part of the ongoing challenge of collecting tax revenue. In his single-country study of Brazil, Weyland (1998) measures the rise and decline of Brazil’s “Developmental State” with taxation revenues. For students of the advanced, industrialized countries, where taxation grew substantially during the 20th century, taxation has been used as a way of studying the degree to which nation-states persist in the wake of presumed pressures towards state diminution or disintegration (Garrett 1998).

### *B. Solving Collective Action Problems*

A second line of research has identified the problem of taxation more squarely as a problem *within* society rather than as merely a struggle between the state and an undifferentiated society. Analysts have asserted that one of the central constraints on the state’s ability to collect is the underlying collective action (or free rider) problems inherent in the demand for taxation (Bates 1989; Levi 1988; Cheibub 1998; Steinmo 1993, 1998). Rather than assuming that governments or states collect taxes for their own alien needs, this line of analysis assumes that citizens want the goods and services that modern states can provide, but that they would rather someone else pay.

The collective action problem is argued to appear at the point of policy-making and/or during the process of collections. In the first case, the problem of taxation involves getting citizens, groups, and/or politicians to agree to a set of policies that will generate tax revenues from within society. Revenue imperatives stand in conflict with temptations on the part of citizens and groups to challenge the imposition of taxes, or to dilute the total burden with tax incentives and loopholes. For analysts such as Steinmo (1998) and Garrett (1998), who do not discuss administration at all, we can assume that they interpret variations in tax collections to reflect variation in the outputs of policy making. In other words, political conflicts are manifest almost exclusively at the level of policy and law.

According to many others, however, the administration of taxes is also a critical and analytically separable component of generating tax revenues. For example, Levi argues that some level of quasi-voluntary compliance is ultimately necessary to generate tax revenue because people tend to find ways to avoid and/or to evade the tax burden (Levi 1988: 49). Particularly when sufficient numbers of people do not accept the state's demands for taxes as legitimate, collections are likely to suffer. From this perspective, significant levels of tax collections imply that a sufficient share of the citizenry has been persuaded to see beyond narrow interests, and to contribute to the collective welfare through tax payment.

### *C. Distributive Justice*

A third central concern of political scientists has been with the distributive dimension of taxation. Analysts have highlighted the observation that resources tend to get allocated in an unequal manner within market economies, and that the state generally plays a role in either exacerbating or ameliorating such inequalities through the differential allocation of the total tax burden. In this light, taxation is not simply relevant in terms of how much revenue is collected by the state from society, but in terms of who pays what. For some scholars, the allocation of the tax burden reflects more broadly on how equity and fairness are defined across countries and over time. Particularly when the analyst provides a baseline notion of what is fair and equitable, the *problem* of taxation is more than simply a collective action problem, but a political contest involving struggles over power and definitions of fairness.

Typically, equity issues are evaluated from one of two angles—either from the perspective of “vertical” or “horizontal” equity. Again, within the economics literature, and in policy studies more generally, such concepts are often treated as unproblematic, while in practice, definitions are highly contested among analysts as well as stakeholders within tax systems. The canon of vertical equity demands that those with greater economic resources should pay more than those with lesser resources as a share of total economic well-being because the “degree of utility of income decreases when income increases” (Cohen-Stuart [1889] 1964: 48). In other words, a “just” tax system, which treats citizens “equally,” should be progressive. The canon of horizontal equity demands that those with *similar* economic resources should carry similar burdens of taxation. When analysts or stakeholders identify violations of such principles within a tax system, they can claim that a tax system is “unfair.” However, political debate is likely to follow over the question of how to measure taxable resources. For example, taxable units may be defined in terms of individuals, households, and/or firms within society, and the measurement of inequality may be based on income, property, level of opportunity, race, marital status, historical legacy, or other factors.

### **III. Measuring Taxation**

Measuring the concepts of state capacity, collective action, and distributive justice using tax revenue data entails multiple choices and tradeoffs. Annual

taxation data are regularly measured and disseminated widely by governments and international organizations, particularly through the IMF's *Government Finance Statistics Yearbook*, the World Bank's *World Development Report* and *World Development Indicators* database, as well as through various OECD publications, which provide dozens of time-varying revenue indicators for over 100 countries in both printed and electronic formats.<sup>3</sup> Since scholars may easily create additional indicators through simple mathematical manipulations, particularly in combination with other economic and demographic indicators, the array of options for measuring taxation becomes rather wide.<sup>4</sup>

The central challenge for comparative taxation scholarship is to develop the best fit between the indicator(s) and the particular concept under investigation. It is worth emphasizing that while the amount of tax revenue collected by a government may be of some intrinsic interest, the central concern here is to consider how tax collections can be used as *indicators* of particular phenomena or processes. To say that a tax indicator has measurement validity implies that different levels of collections will correspond with cross-national differences on the analytic construct. For example, if a measure of total tax collections is used as a proxy for state capacity, countries that collect more taxes should be characterized by greater "capacity" in the ways that the analyst defines. Assumptions about which factors influence taxation outcomes across countries and over time necessarily involve abstractions from reality, and for various reasons the relative scoring of countries may not adequately reflect their relative position on the range of variation on the variable. Making choices about measurement is likely to be an iterative process involving some hypothesis testing and potentially some reformulation of the central concept under consideration (Adcock and Collier 2001: 4-11). Often, this is an implicit thought process, but the goal of this section is to make explicit some of the central trade-offs that must be evaluated in order to construct a valid indicator. I highlight the potential tensions between measurement choices and observed scores by identifying potential challenges to certain assumptions implicit in the study of taxation and by describing how certain "benchmark cases"—cases that are generally well accepted as "cases of" something—actually score on particular indicators.

Reflecting on more than a decade of scholarship on the comparative political economy of taxation, we can identify the tensions and trade-offs associated with measuring taxation, and how such choices force us to consider the conceptual relationship between the state and society. Because there is minimal discussion of measurement and conceptual concerns in these works, it is generally difficult to say with certainty that a particular measurement instrument would have been preferable. In some cases, it was practically impossible to replicate previous analyses with alternative measures, and in other cases, the analyst provides results using multiple indicators. In one case (Cheibub 1998), I reestimated several of the models using alternative measures, and while I found minor differences in the results, it does not appear that the broad substantive findings would have differed dramatically. The point of this analysis is to demonstrate that discussion of measurement concerns aids in our understanding of the particular concepts under investigation, and to flesh out infer-

ences from the various studies based on choices concerning measurement. Such discussion should provide the basis for making measurement choices in future work.

As depicted in Table 1, most tax collection indicators are constructed as ratio variables that relate collections to some standardizing measure. Rather than evaluating any specific indicator, I decompose the process of constructing taxation indicators more generally into three sets of choices or questions, and examine how each of these choices may affect measurement validity:

- A. Which streams of revenue should be included in the indicator? That is, should all tax revenues be included, only specific revenues, or a combination?
- B. What should revenues be measured in relation *to*? Most, though not all, cross-national studies of taxation use ratio variables to serve as taxation indicators because “dollar-denominated” comparisons are generally not meaningful.
- C. How should cases be defined, and what is the relevant universe to which comparisons can be made?

I take up each of these concerns later, discussing the impact of different strategies on the content validation of indicators using conceptual reasoning (Adcock and Collier 2001: 14).

#### *A. The “Numerator”: What Should be in? What Out? And Why?*

The most important step in the construction of a taxation indicator is the process of selecting and justifying which streams of revenue should be included and which should not. These choices are based upon assumptions about incidence (who pays) and theoretically informed insights regarding what types of social and political dynamics affect revenue outcomes. The menu of options is wide, but finite, and it is useful to begin by looking at the range of government revenue sources in order to highlight how analysts can make decisions about which streams of revenue best capture the underlying construct and which may simply add “noise” to the measure. (See Table 2 for a list of the range of the major government revenue streams, as classified by the IMF.)

*1. Distinguish tax from non-tax sources of revenue.* First, it is necessary to identify the sources of government revenue that are *not* taxes. As described earlier, many other forms of government revenue, such as from the sale of postage stamps, school fees, or road tolls, do not fit the definition of taxation. As Cheibub points out, the distinction between tax revenue and the state’s ability to extract revenue from society via other means is a crucial one as taxation is a central path for escaping fiscal crisis (Cheibub 1997: fn 36). Because such revenues are collected with explicit reference to individual benefits, even if some of the “profits” generated from such services may be used for other government services, there is seldom a free rider problem associated with such revenues. Similarly, when a state generates surpluses from its ownership of certain productive firms, this cannot be considered taxation, except possibly when the firm is run along business principles, and income and/or production



**Table 2**  
**Line Items from the Government Finance Statistics Yearbook**

Table A

Revenue and Grants, Consolidated Central Government

I Total Revenue & Grants (II + VII)

II Total Revenue (III + VI)

III Current Revenue (IV + V)

IV Tax Revenue

1. Taxes on income, profits, and capital gains
2. Social Security Contributions
3. Taxes on payroll and workforce
4. Taxes on property
5. Domestic Taxes on Goods
6. Taxes on International Trade and Transactions
7. Other taxes

V Non-tax Revenue

VI Capital Revenue

VII Grants

(Replicated for "State, Region or Province Governments" and "Local Governments")

---

Source: International Monetary Fund (1996), *Government Financial Statistics Yearbook*.

are taxed in the same manner as private firms. Financing the state through monetary instruments—using inflation or money printing to generate additional resources for the state (modern forms of *seigniorage*)—requires no participation on the part of society.

While it is true that certain financing strategies, such as the use of selective credit or subsidies, particularly within an inflationary environment, wind up placing an indirect burden on particular groups, and are casually referred to as "taxes," such instruments have no participatory or compliance component, and ultimately say little about the state's relationship with society. Rather, such forms of finance can be used as part of a benchmark measure with which to evaluate the degree to which states collect "real" taxes (see discussion of the "denominator" below). Opportunities for financing the state in these ways may influence tax systems, and vice versa, but tax revenue remains a unique source of revenue with respect to the challenge it presents for distributing the burden within society and coordinating payment with the state.

Not all analysts agree that non-tax revenues should be excluded, however. In her study of the tax capacity of government, Fauvelle-Aymar argues that it is necessary to employ a measure of government revenue that includes property income, but not capital revenue or grants. She argues that particularly for the study of developing countries, it is desirable to use this broader definition because when firms from the industrial and/or commercial public sector "abuse their monopolistic position, then their profit becomes similar in 'political' terms to taxation" (Fauvelle-Aymar 1999: 402). Moreover, she explains that the reliability of the taxation/revenue indicator improves when non-tax revenues are included because many key industries are nationalized in developing countries, and the omission of such revenues would lead to an underestimation of state capacity (Fauvelle-Aymar 1999: 403). In this case, measurement consid-

erations clearly forced the scholar to reflect carefully on the concept of “capacity.”

Empirically, decisions about measurement prove consequential, as demonstrated by a comparison of results of regression analyses across two models—one with a dependent variable that includes non-tax revenues, and another with a dependent variable that excludes those revenues. Significant differences are found in the impact of mineral wealth, regional dummy variables, and a dummy variable for political opposition (Fauvelle-Aymar 1999: 408-9). Although the author provides explanations for the differences, there is good reason to believe that the alternative measures are actually measuring different concepts, rather than simply triangulating on a single concept. This is particularly clear when considering the impact of the mining sector—which is a robust determinant of the revenue measure that includes non-tax revenue, but not of the measure that excludes such revenues. When mining revenues are included in the taxation indicator, this makes *rentier* states appear stronger in terms of “capacity,” and would lead one to predict that the existence of a mining sector produces higher levels of state capacity, *ceteris paribus*. Given that mineral-based revenues generally require little technical or political capacity to procure, these findings imply a serious threat to the measurement validity of an indicator that includes non-tax revenues.

Moreover, the reporting of non-tax revenues introduces additional sources of measurement error. Government accounts generally only report some, and not all, of these forms of revenue, and the true financial position of the public sector may be hidden across levels of government (see later discussion of units). Often, governments will use extra-budgetary accounts and/or special financing strategies that may obviate the need to rely upon other forms of revenue or deficit financing, but which may or may not reflect a state’s *need* for revenue and inability to raise such funds through taxes. Given the potentially opaque nature of government data on non-tax revenue, this may be an intractable problem, particularly for large-N studies.

*2. Determine which tax revenue streams to include.* Seven major categories of tax revenue—income tax, social security contributions, taxes on payroll and workforce, taxes on property, domestic taxes on goods, taxes on international trade and transactions, and “other” taxes are generally reported by governments in compliance with IMF guidelines. As a result, comparative annual collections data are available for each of these, with further disaggregation for most countries. By briefly describing each of these tax bases, it is possible to make some conjectures about the extent to which one or another revenue stream may adequately serve as an indicator for the three central analytic constructs discussed in this article (these arguments are summarized in Table 3). Each revenue stream varies in terms of incidence, transparency, administrative ease, and its connectedness to particular expenditure outcomes. Different assumptions about the political, administrative, and other demands of the respective tax streams affect the degree to which they may be considered good measures of the broader concepts identified in the previous section. Undoubtedly, a central tension with all of these measures is that the precise ways in which taxes are levied in dif-

ferent countries varies, and the analyst must determine the extent to which such variation ultimately undermines the validity of the collections measure.<sup>5</sup>

*a. Taxes on income, profits, and capital gains and property taxes.* Most analysts would probably agree that the “purest” form of taxation includes those taxes levied on income, profits, and capital gains. Such taxes are paid over to the state directly by individuals and firms, often with graduated rates for different levels of income. Political scientists studying taxation have tended to develop indicators based on the standard assumption that these taxes have the qualities of being among the most progressive, most difficult to administer, most transparent, and least required of any government revenue streams. As a result, such revenues tap into all three of the standard analytic constructs. That is, higher levels of income tax collections are generally associated with greater levels of capacity, collective action, and downward redistribution. In the case of the first two concepts, however, if only upper-income groups are liable for such taxes, collections of this tax will only reflect state capacity and/or collective action with respect to those upper groups.

**Table 3**  
**Evaluating the Measurement Validity of Tax Revenue Streams**  
**as Indicators of Various Aspects of Political Life**

Concept:	State capacity	Collective action	Distributive justice
<b>Revenue stream:</b>			
Income, profits, capital gains taxes	<b>High:</b> Requires extensive monitoring, enforcement.	<b>High:</b> Significant opportunities for individuals and groups to attempt to free ride.	<b>High:</b> Generally thought to be progressive.
Property taxes	<b>High:</b> Requires extensive monitoring, enforcement.	<b>High:</b> Significant opportunities for individuals and groups to attempt to free ride.	<b>High:</b> Generally thought to be progressive.
Domestic consumption taxes	<b>Medium:</b> Requires extensive coverage, but monitoring/enforcement less than income taxes.	<b>Low:</b> Generally hidden and incremental, so minimal free rider problem.	<b>High:</b> Generally thought to be regressive (i.e., more collections implies upward redistribution)
Social Security Contributions/ Payroll and Workforce taxes	<b>Medium:</b> May require substantial administrative capacity for any collections.	<b>Low:</b> Benefits generally tied to contribution, so minimal free rider problem.	<b>Low:</b> Benefits generally tied to contribution; incidence more ambiguous.
Other taxes	<b>Low-medium:</b> Capacity requirements are ambiguous, but still requires <i>some</i> capacity.	<b>Low:</b> Ambiguous relationship to free rider problem.	<b>Low:</b> Incidence is ambiguous.
International trade and transactions taxes	<b>Low:</b> Requirements tend to be more limited than for most domestic bases.	<b>Low:</b> Policy making and administration do not present free-rider problems similar to domestic tax bases.	<b>Medium:</b> Import taxes generally thought to be regressive, but luxury items may be taxed more heavily.

Each cell evaluates the likely measurement validity of the revenue stream with respect to the corresponding concepts in the columns of the table. Evaluations are based on conceptual reasoning, assuming comparisons across countries at all levels of development.

The incidence assumptions with respect to income taxes are not without challenge, however, as some scholars have argued that income taxes levied on firms may often get passed on to consumers, and that high-income individuals may be the biggest evaders of personal income tax, making this source of revenue regressive (Shah and Whalley 1991: 178-9). Nevertheless, *perceptions* of tax incidence may be at least as important as actual incidence given certain theoretical justifications for studying taxation (Webber and Wildavsky 1986: 522). Moreover, when it comes to questions of capacity and collective action, collections of taxes on income, profits, and capital gains still reflect levels of state-society and intra-society coordination and cooperation. As a result, it is still reasonable to conclude that this revenue stream is a solid indicator of all of the three central constructs. Case-by-case analyses may be useful to determine the extent to which the assumption of progressivity is violated.

Taxes on property generally exhibit the same characteristics as taxes on income, profits, and capital gains, and in fact, the World Bank simply reports these two revenue streams together as an undifferentiated sum in its development indicators dataset. In many ways, property taxes may be a far better measure than income taxes because there are fewer debates about incidence. However, because this tax plays such a small role in the financing of virtually all central governments, it cannot replace the more important income tax measure.

In practice, the combined measure of income and property taxes appears to generate solid results in comparative analyses of key concepts. We do not have tried and true measures with which to conduct rigorous evaluations of how this (and other) revenue streams perform as measures, but it is possible to consider if the measure produces scores that we would “expect” for benchmark cases. In table 4, I carry out this exercise, comparing how a select group of countries score relative to 91 countries (those with available data) in terms of income, profits, capital gains, and property taxes as share of GDP. In this table, each of the eight countries are placed in either the top or bottom row, depending upon whether we should *expect* them to score low or high based on our prior knowledge of their state capacity, inclinations towards collective action, and the state’s role in redistribution. Then, each country is placed in either the left or right column based on whether their *actual* rank score is low or high. Countries such as Denmark, the United Kingdom, and Australia are all generally recognized as having well-developed central states, in which the state authority generally manages to elicit compliance from its citizens, and in which the state certainly plays some role in redistributing resources downwards. As predicted, all of these countries score within the top five of all countries in terms of collections. Moreover, countries such as Peru, Rwanda, and Madagascar, which are widely recognized as having much less developed states with less effective authority to elicit collective action within society and tend to be unable to redistribute significant resources downwards, score, as expected, very low on the taxation indicator. While most countries for which we might have prior expectations score in the range expected, there were some surprises. For instance, within Latin America, Uruguay is considered one of the most well-developed states, characterized by high levels of citizenship, and where the

**Table 4**  
**Expected and Actual Country Tax Indicator Scores**  
**(Income, Profits, and Capital Gains Taxes/GDP)**

Expected Score: (Based on country knowledge)	Actual Score (Rank out of 91)	
	LOW	HIGH
LOW	<p style="text-align: center;">✓</p> <p style="text-align: center;">(Actual score similar to "predicted" score)</p> <p style="text-align: center;">Peru (88) Rwanda (75) Madagascar (81)</p>	<p style="text-align: center;">✗</p> <p style="text-align: center;">(Actual score <i>not</i> similar to "predicted" score)</p> <p style="text-align: center;">Venezuela (5)</p>
HIGH	<p style="text-align: center;">✗</p> <p style="text-align: center;">(Actual score <i>not</i> similar to "predicted" score)</p> <p style="text-align: center;">Uruguay (78)</p>	<p style="text-align: center;">✓</p> <p style="text-align: center;">(Actual score similar to "predicted" score)</p> <p style="text-align: center;">Denmark (9) United Kingdom (8) Australia (3)</p>

Country rank order is indicated in parentheses. Rankings are based on average (1990-94) income and property taxes as share of GDP.

Source: World Bank (1998).

state plays an important redistributive role, whereas Venezuela is often considered a relatively weak state, where corruption is thought to be high, and where extreme inequalities persist. Nevertheless, Venezuela scores fifth (above Denmark) and Uruguay 78th (below Rwanda) of the 91 countries measured. What are we to do with such discrepancies?

When scholars identify such mismatches between preconception and actual measure, they are ultimately faced with four choices: They can (1) trust the measure to reclassify the case in their own mind—that is, to alter their characterizations of the countries; (2) question the measurement validity of the indicator, electing to choose an alternative; (3) reframe the construct in light of the indicator scores; or (4) maintain their understanding of the case(s), while continuing to use the indicator and the original concept, and to declare that the mismatch is simply the product of unexplained "noise." Depending upon the nature of the study and the number of cases involved, any of the four solutions could be viable, but the reader benefits greatly when the author highlights such tensions and is clear about how such questions of measurement validity are addressed. It is probably fair to hypothesize that Venezuela's high score is due to a special type of income tax collection (oil revenues classified as income taxes) and for particular historical reasons, Uruguay's low score can be understood from the perspective that policy makers have opted to redistribute in-

come on the expenditure side and to collect taxes through more indirect means. Such insights potentially contribute to the reframing of the construct, or they may provide insights for a causal argument. Alternatively, the particular scores and historical insights may be deemed irrelevant and epiphenomenal. They may even serve to make analysts rethink their preconceived characterizations of the two cases.

*b. Social security contributions/payroll and workforce taxes.* Social security contributions are monies paid over by citizens to the state with the express purpose of funding particular social expenditures such as pensions, unemployment benefits, or health care. Governments and international organizations often list separate line items for payroll and workforce taxes, but these are generally similar types of revenues in terms of administrative requirements and incidence assumptions. While such contributions are described as taxes in everyday discourse, and for individual citizens/taxpayers they may seem quite similar to income taxes that may also get withheld by employers, the fact that they are usually intended to be contributions to funds that pay out only to contributors at a later date suggests that one could make the case that they should not be classified as taxes, at least according to the strict definition identified earlier. Alternatively, they can be identified as a form of regressive and earmarked financing associated with quite different types of political imperatives than is the case with other forms of general taxation, as Steinmo describes when investigating the introduction of this tax in the U.S. (Steinmo 1993: 99). Indeed, because social security revenues are tied to particular plans, which vary quite widely across countries,<sup>6</sup> assumptions about the political ramifications of (non) collection may need to be exercised with more caution than for other streams of revenue.

Particularly for scholars interested in measuring the sheer scope of the state, a finding of significant differences in the collections of this form of revenue does indicate differences in the size and efficacy of the state, *ceteris paribus*, but social security and payroll taxes do seem less valuable for tapping into the other constructs discussed here. In his study of Brazilian state capacity, Weyland defends his decision to exclude such revenues from his taxation indicator "because the government cannot use these resources as freely as tax revenues," and because "their quasi-contractual character makes it politically easier to raise these sources of revenue" (Weyland 1998: 69, fn 7). In other words, because of the earmarked nature of the contributions, which can resemble market-based transactions in the sense that costs and benefits are so closely tied, free riderism is much less problematic than with other forms of revenue. That is, collections do not really reflect collective action *per se*. Moreover, the distributive implications of such revenues are not entirely clear (Shah and Whalley 1991: 183). Because benefits may be so closely linked to contributions, the net impact may be perceived to be flat.

*c. Domestic consumption taxes.* Domestic consumption taxes, generally levied as sales, excise, or value added taxes (VAT), are largely paid by retailers, service providers, and manufacturers as a share of the value or quantity of goods

sold. Consumers ultimately carry the burden of such taxes, paying higher prices for goods and services. As an indicator of larger political dynamics, the collection of such taxes tends to be used most prominently as a measure of distributive justice. Consumption taxes are generally thought to be regressive, as the proportion of income consumed by poorer households tends to be much greater than for wealthier ones. As a result, this revenue is a useful indicator of regressivity in the state budget. Again, there are important exceptions to this incidence assumption, as particular countries may have graduated consumption tax rates, placing a higher burden on luxury items, and lower or even zero-ratings on basic goods that may be consumed disproportionately by the poor.

In terms of the other two concepts, domestic consumption taxes are likely to be somewhat more ambiguous as measures. Collection of consumption taxes can be interpreted as evidence of a functioning and competent tax administration, but not to the degree that is generally associated with the collection of income taxes. Taxes on consumption still generally require significant bureaucratic capacity, but not nearly the same amount of information is required as with the taxation of income, and these revenues tend to be easier to collect than taxes on income because they are collected indirectly, incrementally, and generally at the point of purchase. Consumption tax collections are also probably less useful as measures of collective action. Because such taxes tend to be more hidden, and particularly through the VAT, they are designed to be self-enforcing within society, the free rider problem is less severe, and thus collections cannot be interpreted as a *strong* demonstration of a society solving a collective action problem.

*d. Taxes on international trade and transactions.* Taxes on international trade and transactions are generally paid over to the state by the agents transporting commercial goods (or services) over national borders. In certain ways, import taxes exhibit the same properties as consumption taxes in the sense that they tend to fall on consumers in the form of marked-up prices on goods and services. As a result, such taxes may be similarly useful as indicators of distributive outcomes, in the sense that both taxes tend to be regressive. However, it is difficult to interpret collections of this form of revenue with respect to the other two constructs. Strategies for developing domestic industries have often motivated the imposition of trade taxes to a much greater degree than the need for revenue. The fact that most state bureaucracies make a clear distinction between “inland revenue” and “customs” further reveals the extent to which these are potentially different processes. Because they tend to be collected at ports and on narrow bands of the economy, the types of capacities and political compliance necessary for collection are rather different than those associated with domestic tax bases. The well-established inverse correlation between level of development and trade taxation underlines the assertion that the level of trade tax collections is probably not a very good predictor of state development or capacity across country cases at varied levels of development.

*e. Other taxes.* Finally, the commonly used line item, “other” taxes, is a revenue stream that needs to be reconciled in conceptual terms. Such taxes tend to

get levied in nationally distinctive ways to the extent that they do not fall under any of the other broad categories. Yet, these revenues still qualify as taxes according to governments and/or international organizations. For example, stamp taxes or licenses that must be paid during the course of certain transactions are generally clustered in the "other" column in national accounts. While the implications of collecting such taxes for understanding distributive justice and collective action outcomes are ambiguous, such revenues are likely to shed some light on capacity. That is, the exclusion of such revenues from the indicator might result in an underestimation of the capacity of a country that manages to collect a significant amount of tax revenue through unconventional means. In practice, however, the impact is not particularly significant, as in 1990, only one country (Romania) collected more than four percent of GDP in such taxes, and the average collections was less than one percent of GDP. Moreover, there is not significant variation within this category to heavily influence the relative scoring of countries.

*3. Options for combining revenue streams.* All of the studies considered in this article use at least one indicator that combines two or more of the earlier-mentioned revenue streams because various forms of tax collection do appear to overlap as indicators of several political processes. Because taxation indicators are themselves related, and real-world decisions to tax are almost always made with respect to other sources of revenue, it may be absolutely necessary to consider various forms of tax revenue together. Decisions about how to combine revenue streams in the measurement instrument should be based on evaluations of possible threats to measurement validity.

*a. Total taxation.* The most obvious and widely used revenue combination is "total tax revenue," which political scientists and other scholars often interpret as representing the overall fiscal effort of society or the size or capacity of the state, depending upon the conceptual framing of the project. When scholars rely solely on this measure (Cheibub 1998; Steinmo 1998), this implies that incidence is much less important than the overall burden on society as a single, collective actor. Given the complications of measuring tax incidence, the choice to use total tax collections may be a reasonable strategy, but aggregation does not necessarily solve all of the analyst's problems.

To evaluate the measurement validity of a particular formulation of "total taxes," we can compare scores with and without certain revenue components. For example, if social security contributions do not measure capacity in the same way that income taxes do, this can be a source of measurement error, but not a dramatic one. Looking at 58 countries for 1990 and comparing rank orders with and without social security revenues revealed an average difference in rank of 4.6, and seven countries had rank order differences greater than 10.<sup>7</sup> As shown in Table 5, in a time series cross-sectional dataset of 135 countries for the period 1970-1990, the correlation between indicators including and excluding social security revenues is very strong for all countries ( $R = .94$ ). However, the relationship between the standard (all-inclusive) total tax measure and one that excludes trade revenues is weaker ( $R = .77$ ). As will be later



**Table 5**  
**Bivariate Correlations of Various "Total Tax" Indicators (1970-1990)**

Measure:		TOTAL1 Total Taxes	TOTAL2 Total Taxes (less trade duties)	TOTAL3 Total Taxes (less social security)	TOTAL4 Total Taxes (less trade duties and social security)
Universe of cases:					
All countries (n=932)	TOTAL1	1.00			
	TOTAL2	0.77	1.00		
	TOTAL3	0.94	0.75	1.00	
	TOTAL4	0.75	0.97	0.80	1.00
Low- and Middle-income countries (n=803)	TOTAL1	1.00			
	TOTAL2	0.69	1.00		
	TOTAL3	0.93	0.71	1.00	
	TOTAL4	0.65	0.96	0.75	1.00
High-income countries (n=129)	TOTAL1	1.00			
	TOTAL2	0.97	1.00		
	TOTAL3	0.85	0.77	1.00	
	TOTAL4	0.90	0.86	.97	1.00

discussed, the implications of the particular choice to include or to exclude certain revenue streams varies according to how the universe of cases is defined.

*b. Direct vs. indirect taxes.* A slightly more nuanced aggregation strategy involves differentiating between "direct" taxes, which are collected from those citizens who actually pay the tax burden themselves, and "indirect" taxes, which are collected by intermediaries such as producers or firms, but ultimately affect the after-tax income of other actors within society such as consumers and workers. These distinctions tend to be helpful in providing measures of our central analytic constructs, as increased collections of direct taxation tends to reflect significant capacity, overcoming the free rider problem, and increased progressivity. For example, in her study, Chaudhry (1997: 201) uses direct taxation to indicate that certain types of collections require more state capacity than others. By contrast, high collections of indirect taxation are interpreted as reflecting regressivity in the tax system, while being more ambiguous with respect to capacity and collective action.

More generally, the challenge for scholars is to make decisions about which revenue streams should be classified as direct and which indirect. Here, income and property taxes are almost always regarded as direct, and domestic consumption and trade taxes are almost always considered indirect, but it is less clear how to classify social security, payroll, and "other" taxes. Using the observations about the inferential possibilities and problems discussed earlier with respect to these revenue streams, analysts should make choices about which

aggregations would best represent the concept under investigation, rather than complacently accepting “standard” practices of classifying direct and indirect taxes.

### *B. The “Denominator”: Converting Currencies into Meaningful Units*

Although governments ultimately collect their taxes in local currencies and report levels of collections in these units, in this form, such measures are generally not very useful to scholars making cross-national comparisons. In order to carry out comparative analysis it is necessary to “standardize” or “normalize” by constructing a ratio variable.<sup>8</sup> Even converting the collections revenue streams into a standard currency such as U.S. dollars or international purchasing power parity units is insufficient. The fact that the U.S. may collect X trillion international dollars of tax revenues and another country collects only one-fifth that amount is only meaningful in a comparative sense in relation to the size of the population or the size of the economy of the respective countries.

In studies of taxation, analysts generally opt for one of three strategies with respect to the problem of non-homogeneity of currency units: (1) they may report collections as a proportion of national income; (2) they may report collections as a proportion of total taxes or total revenues; or (3) they may simply opt against using a reference or standardizing variable and report tax collections in local currency and use other contextual interpretations of the revenue statistics. In any of these cases, the analyst may look at one or more country cases over time and report rates of change. Predictably, there are advantages and disadvantages to each of these strategies.

*1. National income measures.* By far the most prevalent measure used to standardize tax collection revenue indicators across time and space is national income. Most analysts report tax revenues as a proportion of Gross Domestic Product (GDP), or sometimes as a proportion of Gross National Product (GNP). Generally speaking, this strategy is implemented because analysts implicitly agree that the problem or challenge for states is to collect a portion of the total economy in tax revenues, and that opportunities and constraints on taxation for policy makers and the bureaucracy are ultimately determined by the size of a country’s economic output. These are quite reasonable assumptions, and provide enormous leverage for cross-national comparison in a world in which the size and wealth of countries are radically different.

Yet, it is critically important to recognize that even GDP is itself only an *indicator* of the size of the annual output of the economy and is an imperfect proxy for the availability of tax handles within a given society.<sup>9</sup> If they were available, other measures of societal income or wealth might serve as better normalizing indicators given a particular conceptual understanding of taxable resources. Moreover, it may be incorrect to assume that the challenge of tax collection on the part of the state is proportional to the size of the tax base. When GDP is used as the denominator in the tax indicator, the implication is that a country with a GDP of Y dollars would need to collect twice as much in

tax revenues in dollar terms to score identically on the tax collection indicator as a country with a GDP of  $0.5*Y$  dollars. Depending on the size and distribution of the population and of wealth within the society, this may or may not be a valid assumption. For example, if one person controlled all the wealth within society, that would clearly pose a much different challenge for capacity or collective action than if income and wealth were perfectly distributed.

2. *Total revenue.* A second measure often found in the denominators of taxation indicators is total revenue or total taxes. When one of these measures is used, the implication is that the *structure* of revenues is meaningful and that certain revenue streams should be interpreted differently from others, perhaps as discussed earlier. For example, one might measure income taxes in proportion to total revenues as a measure of progressivity (Steinmo 1993: 3) or direct taxes as a share of total revenues as a measure of state development (Chaudhry 1997: 201). The advantage of such a strategy is that it highlights the trade-offs available to states and societies in how the state can be financed, focusing our attention on the interpretation of revenue structure “choices.” For example, Weyland measures income taxes in relation to both total tax revenues and total “fiscal” revenues (which include non-tax “parafiscal” revenues) in order to characterize the degree to which progressive taxes have been a component of state finance in Brazil (Weyland 1996: 124). The implication is that the choice to finance the state through one form of revenue over another reflects low or high scores on the underlying concept (i.e., state capacity, collective action, or redistributive social policy).

Unfortunately, this strategy also has drawbacks. First, as described earlier, the analyst may not be able to measure non-tax sources of finance reliably because many of these are quite hidden and not reported in national accounts data. Moreover, because countries do collect different levels of taxation and revenue overall—for example in relation to national income—the relevance of the structure of taxation alone may not be very telling. For example, both Finland and Sierra Leone collected 31 percent of total current revenue in income and property taxes in 1990. Generally speaking, we do not think of these two countries as being very similar in terms of state-society relations, suggesting that we should either rethink our measures or perhaps rethink our understanding of these cases. Yet, when we find that Finland collects more than seven times as much revenue as share of GDP than does Sierra Leone, similar scores on the income and property tax as share of GDP measure make more sense. Nonetheless, the question remains, should any measure that scores these two cases as equals be considered a valid measure? Again, this depends. For comparing countries with widely different total tax burdens, the answer is generally no. For countries in which the total magnitude of the burden differs (but not dramatically or to a degree that influences our understanding of the cases) revenue structure components may provide useful and valid measures.

3. *No denominator.* Some analysts reject the idea of using a “denominator” altogether, and simply report collections in local currency. In such cases, the amount of tax collected must be interpreted by the author with reference to

specific country knowledge, particularly with respect to a particular moment in time, and the unstandardized revenue statistics must be described as representing either a “little” or a “lot” of the particular concept under investigation. For deeply historical studies, there may not be readily available measures of national income or total revenues, or those measures may be considered so unreliable as to be meaningless. In such cases—see, for example, Levi’s (1988) treatment of various fiscal histories and Chaudhry’s (1997) descriptions of early state building in the Middle East—tax revenues are reported in local currencies and with dynamic characterizations regarding over-time growth in revenues. Although there may be good reason to follow such practice (in the absence of standardizing measures), we are left to “trust” the respective analysts to make valid inferences. Replication and verification become extraordinarily difficult for all but the country specialist. On the other hand, for scholars wanting to explore particular cases, such measurement strategies may be the only option.

### *C. Fine-Tuning*

The simple *availability* of taxation data does not imply that all countries and cases can be readily compared in a meaningful way. More careful measurement strategies are often necessary to make valid inferences. Otherwise, the inclusion of non-comparable cases in cross-national studies of taxation can aggravate the problem of “conceptual stretching” (Sartori 1970). To avert such problems, taxation indicators can be “fine-tuned” with clear specifications of the unit of analysis, the universe for comparison, and the scale of measurement.

*1. Unit of analysis: Determining which levels of government constitute the “state.”* A first central concern is with respect to the standardization of levels of government and the notion of the “state.” Here, as is often the case, measurement issues may force the analyst to reconsider conceptual problems. Specifically, scholars must be clear about whether the concept of the state includes only the central government or the entire public sector, including provincial/state and local/municipal governments. Even in countries that are not generally recognized as federal, subnational units may collect some taxes. Fortunately, public accounts data almost always make clear distinctions between tax revenues (and other revenues and expenditures) that are raised exclusively at the central state levels and those raised at other levels of government. Typically, when revenues from all levels of government are included together, this is identified as “General” government revenue. Depending upon the analyst’s theoretical interest, the choice of how to define the state is critical, as the share of subnational tax revenues as a proportion of total tax revenues varies widely across countries.

On the one hand, it stands to reason that in attempting to measure the size of the *national* state and its ability to collect tax revenue for a general public, only central state collections should be used. At lower levels of government, particularly the local or municipal level, the political challenges associated

with collecting taxes is rather distinctive. The payment of central government tax obligations is distinct from payments made within the market because it is far more difficult to link private benefits to the costs of payment. Yet, at more local levels of government, it stands to reason that interests, needs, and resources are more homogeneous, mitigating the more generalized collective action problem that may exist at the national level. This implies that collections across levels of government may need to be interpreted as solutions to different types of political problems. Similarly, the size and/or capacity of the “state” may not be well reflected when provincial and/or local revenues are included, because, for example, such authorities may predate the very existence of the national state, implying that collections are potentially negative indicators of the relative success of a state-building process. In fact, one important strategy for measuring the consolidation of central state authority has been to compare over-time changes in collections at the national and local levels (e.g., Levi 1988: 150).

Because some analysts conceive of the state and/or interpret collections differently, they opt not to exclude subnational revenues from their indicators. They would argue that such revenues may strongly influence and be influenced by many of the same processes that affect national-level taxation, and for the purposes of analysis they may really reflect a part of the same outcome. In other words, omitting them would imply an *underestimation* of total tax capacity or size of the tax state. The challenge of taxing citizens who are already heavily taxed by local or provincial governments is likely much more difficult than those who are only lightly taxed or not taxed at all, and decisions about national taxation may be coordinated with subnational tax efforts, suggesting that such revenues should be included as a source of control. In some cases, central governments may actually collect subnational government tax revenues, or vice versa, potentially making it difficult to make inferences about capacity from national accounts data alone. Citizens may or may not distinguish among levels of government when they pay their taxes, and they may have little information about how such revenues get transferred once they have been collected—suggesting that on average, there is simply a total tax “effort” on the part of society that should be recognized, regardless of where such taxes are paid and collected.

Scholars can address such issues by stipulating assumptions about whether or not the problem of taxation at subnational levels is really part of the same problem as taxation at more local levels, justify that assumption, and develop indicators from that perspective. Steinmo, for example, is centrally concerned with national decision making, but in order to compare tax burdens between the U.S. (a federal country) and Sweden (a unitary state), he compares U.S. taxes plus New York State taxes to the overall British tax burden (Steinmo 1993: 118). Cheibub’s (1998) study of extractive capacity and Garrett’s (1998) of macroeconomic policy appear to use measures that include all levels of government, though this is not explicitly stated or discussed. Interestingly, the political arguments made in these studies are explicitly about *national* politics, and yet given the important degree to which subnational governments *do* collect in many countries, it would seem important to know whether or not the same factors that influence central state collections also affect those collec-

tions. Peters' (1991) study stands out in its presentation of extensive comparisons with respect to the extent to which subnational governments collect, making the important point that within the advanced industrialized countries, subnational revenues in both federal and unitary states represented more than 25 percent of all revenues. On the other hand, his study does not test a particular hypothesis about the determinants of tax structure or how to interpret tax collections within a broader conceptual framework, and we are left with little basis for evaluating what to make of cross-national differences in the degree to which subnational governments collect taxes.

*2. Determining which cases to compare.* A further challenge for comparative taxation analysis is the question of which countries and time periods can and cannot be compared within a single framework or with a single indicator. The determination of the universe of cases to which the taxation indicator applies may be driven by scholarly interests—i.e., whether the analyst is interested in small- or large-scale magnitudes of variation, as well as by assumptions or conjectures about context (Adcock and Collier 2001: 8-12). Again, it is worth reiterating that the goal of constructing the taxation indicator is to develop a metric in which scores are consistently meaningful and if the score in one case cannot be interpreted in the same way for another case, then the measure is invalid, and either it must be adjusted or the universe of cases must be more narrowly defined.

The universe of cases may be delimited along several possible dimensions. The most prevalent strategy is in terms of economic structure, particularly level of development. For example, by studying only the advanced industrialized countries (Steinmo 1993, 1998; Garrett 1998; Peters 1991), the problem of non-comparability with poorer countries may be averted. For studies of the wealthier countries, it is much more reasonable to conceptualize the problem of taxation as one of policy making than of administration, which tends to be the more prescient concern for collection in the poorer countries. Moreover, the prevalence of various tax instruments tends to be correlated with level of development. For example, between 1990 and 1994 social security revenues represented approximately 27.5 percent of total current revenues for the upper-income countries, but only 9.8 percent of total current revenues for upper-middle-income countries, and were negligible in virtually all poorer countries. As a result (as demonstrated in Table 5), the correlation between measures of total taxation that variously include and exclude social security revenues are stronger for the low- and middle-income countries than for the high-income countries. In other words, choices about whether to include or to exclude such revenues in the taxation indicator are more consequential for studies of the high-income countries. Alternatively, the reverse can be said with respect to decisions about whether to include trade duties—a revenue source that is more important in the poorer group of countries. When the universe of cases is defined across levels of economic development, these differences need to be considered in the analysis.

Other strategies for defining the relevant universe of cases provide other sources of analytic leverage. For example, the analysis of cases within a single world region (Chaudhry 1997) may provide a degree of standardization be-

cause economic activities and policy initiatives are likely to be similar across neighboring countries, facilitating the interpretation of revenue data. Other possibilities include limiting the analysis to countries of a particular range of population sizes, to countries with certain types of economic structures (i.e., diversified vs. primary commodity exporters), or to countries with certain legacies of economic institutions (i.e., capitalist vs. communist countries). In each case, scholars can justify decisions regarding inclusion or exclusion of groups of cases based on theoretical or empirical arguments about why revenue data is not likely to reflect the aspect of state-society relations under analysis in a certain country or group of countries.

Almost all studies of taxation identified in this article have been centrally concerned with dynamics and comparisons across time, and it is necessary to consider the measurement problems that can arise in such analyses. For example, in the year 1900 a country collecting 10 percent of GDP in total taxes would be understood as collecting “a lot,” whereas this score would be interpreted as just a “little” in the year 2000. Moreover, the actual content of “capacity” or other concepts being measured varies widely across large expanses of time. In this sense, scholars of comparative taxation find themselves in a similar predicament faced by scholars of comparative democratization who must score cases across historical periods in which context and norms vary widely. Solutions developed by democratization scholars are likely to be helpful—namely the specification of time-specific rules for making inferences from particular measures (Collier 1999: 24-32).

*3. Developing meaningful calibrations and cutpoints for measures.* Finally, a set of issues that should be highlighted when measuring taxation concerns the calibration of meaningful scales from which we can draw inferences. Again, even assuming that our measures are reliable, the plethora of taxation data available should not be “overly” interpreted. An extreme example may be helpful here: if a country’s tax collections drop off by two-thirds between December and January, should this be interpreted as a collapsing state? Or simply the fact that Christmas sales bring in extraordinary amounts of sales tax revenues? In most cases, the proper interpretation is obvious, but this points to a larger issue: for the large-scale, macro-level processes that are being measured by political scientists conducting cross-national studies of taxation, fine-grained data must be approached with caution or they may actually lead to faulty inferences.<sup>10</sup>

The other important calibration issue concerns the age-old issue of the trade-offs between quantification and classification (Sartori 1970) and, in particular, assessing the degree to which interval data can really be interpreted as reflecting evenly spaced differences in scores on the variable. Charles Ragin (2000) aptly describes an important measurement issue when he highlights the existence of “fuzzy-set” problems. For example, in the case of taxation, the differences between two pairs of countries, one comprised of cases collecting 0 percent (Somalia) and 10 percent (Paraguay) of GDP in taxes, and the other collecting 30 percent (Spain) and 40 percent (Italy) of GDP are enormous. Here, the first pair may be interpreted as comprised of qualitatively different country cases—one with a viable state capacity to collect and the other with

none. The differences within the second pair can be interpreted as differences of degree, as both countries collect significant shares of taxation when compared with most other countries. These qualitative differences are not recognized by statistical analyses that would otherwise generate results based on the assumption that a cross-national difference of 10 percent of GDP is homogeneous across cases. One solution is to truncate the universe of cases, so as to investigate only variations of degree, not kind. A second is to recalibrate the revenue data into a new measure that has greater measurement validity based on theoretical grounds. A third is to engage in a combined strategy that recognizes qualitative differences and presents these alongside variations of degree using interval data (Ragin 2000).

#### IV. Conclusion

Tax collection data can be a powerful resource for measuring and comparing state-society relations across countries and over time, but as discussed earlier, numerous pitfalls can lead to the making of faulty inferences from such data. Because scholars choose to study taxation for different conceptual purposes, it is not possible to identify a single “best” measure. Rather, this article has tried to highlight the trade-offs associated with different choices regarding the construction of suitable indicators for specific theoretical constructs. While the strategy of using multiple indicators is a viable one, scholars still need to be attentive to how different indicators may be tapping different analytic constructs, as well as the different types of measurement errors associated with various choices. Too often tax measures are presented as unproblematic indicators of an “important” outcome, when, in fact, the expressly political problem of taxation can be defined in many ways. More explicit discussion of how and why political scientists measure tax revenues would lead to greater conceptual clarity and more powerful theoretical conclusions about the determinants of varied state-society relations.

A central message contained in this article is that scholars interested in the comparative study of taxation should not take for granted the availability of measures and indicators. Such a finding is likely to be useful for analysts using other forms of budgetary data, including expenditure data. The sorting out of conceptual and measurement issues is not merely a tedious “pre-research” exercise on the path to the “real” work of generating and estimating causal models. Rather, the process of specifying and justifying the measurement instrument can be integral to the development of causal theories.

#### Notes

- \* I would like to thank Christopher Achen, David Collier, Marc Morjé Howard, Lucan Way, members of the Robert Wood Johnson Policy Scholars Seminar at Yale University, three anonymous reviewers, and the editors at *Studies in Comparative International Development* for their valuable comments and suggestions.
1. This shortcoming is also evident in studies that use other forms of budgetary data as indicators of larger outcomes. In an otherwise excellent volume on the methodology of studying the



- political economy of the welfare state (Janoski and Hicks 1994), there is no serious discussion of the problem of measurement validity associated with budgetary data.
2. See MacPhail (1998) for an insightful discussion of the lack of attention to conceptual and validity issues in economics.
  3. Indeed, the reliability of government statistics should be questioned, particularly in poorer countries, but these concerns are less tractable and I consider only validity issues here. In other words, I proceed with the heroic assumption that all of the indicators used here are *reliable*. Although the substantive implications of the limits to this assumption are important, they are analytically distinct from an investigation of the extent to which these are *valid* indicators of other ideas or constructs.
  4. Other measures—such as tax rates or more qualitative descriptions of how tax systems work—may be more valid and useful for certain research goals, particularly in studies involving small numbers of country cases, but these may be very difficult to obtain, and time varying data availability varies widely across countries.
  5. In particular, there are significant debates about how to make inferences about incidence from tax structure. See Shah and Whalley (1991).
  6. For example, the political implications of collecting revenues for a system with individual accounts are likely to be quite different than those associated with a “pay as you go” system. Thanks to Lucan Way for making this point.
  7. This analysis was conducted using the same data as presented in Table 4. The sample size is lower because of missing social security revenue data for several cases.
  8. For discussions of the use of ratio variables in comparative analysis, see Berry (1986), Firebaugh and Gibbs (1985), and Kritzer (1990).
  9. In particular, see the discussion of the introduction of GNP as a social indicator advanced by Block and Burns (1986). Beyond conceptual problems, which are central to this article, it is also worth highlighting that there are likely to be serious reliability problems with measures of national income particularly outside of the advanced, industrialized countries. Much of the problem with measurement is related to the question of the informal sector, which tends to be taxed only indirectly, if at all.
  10. Using seasonal models, it may be possible to control for such monthly changes and to develop useful estimates of time series dynamics that *do* reflect the concept under investigation. Comparativists generally find that it is difficult to estimate such models, however, because such seasonal data is rarely available for all of the explanatory and control variables that are of interest. As a result, most time-varying cross-sectional models are estimated using annual data.

## References

- Adcock, Robert and David Collier. 2001. “Measurement Validity: A Shared Standard for Qualitative and Quantitative Research.” *American Political Science Review* 95 (Forthcoming).
- Bates, Robert. 1989. “A Political Scientist Looks at Tax Reform.” Pp. 473-491 in *Tax Reform in Developing Countries*, ed. M. Gillis. Durham: Duke University Press.
- Berry, William Dale. 1986. “Testing Budgetary Theories with Budgetary Data: Assessing the Risks.” *American Journal of Political Science* 30: 597-627.
- Block, Fred and Gene A. Burns. 1986. “Productivity as a Social Problem: The Uses and Misuses of Social Indicators.” *American Sociological Review* 51: 767-780.
- Bollen, Kenneth. 1993. “Liberal Democracy: Validity and Method Factors in Cross-National Measures.” *American Journal of Political Science* 37: 1207-1230.
- Bollen, Kenneth A. 1980. “Issues in the Comparative Measurement of Political Democracy.” *Annual Sociological Review* 45: 370-390.
- Bollen, Kenneth A, Barbara Entwisle, and Arthur S. Alderson. 1993. “Macrocomparative Research Methods.” *Annual Review of Sociology* 19: 321-351.
- Campbell, John L. 1993. “The State and Fiscal Sociology.” *Annual Review of Sociology* 19: 163-185.
- Carmines, Edward G. and Richard A. Zeller. 1979. *Reliability and Validity Assessment*. Beverly Hills: Sage.

- Chaudhry, Kiren Aziz. 1997. *The Price of Wealth: Economies and Institutions in the Middle East*. Ithaca: Cornell University Press.
- Cheibub, José Antonio. 1998. "Political Regimes and the Extractive Capacity of Governments: Taxation in Democracies and Dictatorships." *World Politics* 50: 349-376.
- Cohen Stuart, Arnold Jacob. [1889] 1964. "On Progressive Taxation." Pp. 48-71 in *Classics in the Theory of Public Finance*, eds. R. Musgrave and A. Peacock. London: Macmillan & Co., Ltd.
- Collier, David and Robert Adcock. 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts." *Annual Review of Political Science* 2: 537-565.
- Collier, Ruth Berins. 1999. *Paths Towards Democracy: The Working Class and Elites in Western Europe and Latin America*. Cambridge: Cambridge University Press.
- Dorrington, J. C. 1979. "Kakwani's Measure of Tax Progressivity: A Comment." *The Economic Journal* 89: 652.
- Fauvelle-Aymar, Christine. 1999. "The Political and Tax Capacity of Government in Developing Countries." *Kyklos* 52: 391-413.
- Firebaugh, Glenn and Jack P. Gibbs. 1985. "User's Guide to Ratio Variables." *American Sociological Review* 50(5): 713-722.
- Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy*. New York: Cambridge University Press.
- Goldscheid, Rudolf. [1925] 1964. "A Sociological Approach to Problems of Public Finance." Pp. 202-213 in *Classics in the Theory of Public Finance*, eds. R. Musgrave and A. Peacock. London: Macmillan & Co., Ltd.
- Hyman, David N. 1992. *Public Finance: A Contemporary Application of Theory to Policy*. Chicago: The Dryden Press.
- International Monetary Fund. 1996. *Government Finance Statistics Yearbook*. Washington, DC: International Monetary Fund.
- Janoski, Thomas and Alexander M. Hicks. 1994. *The Comparative Political Economy of the Welfare State*. New York: Cambridge University Press.
- Jones, Charles O. 1974. "Doing Before Knowing: Concept Development in Political Research." *American Journal of Political Science* 18: 215-228.
- Kirk, Jerome and Marc L. Miller. 1986. *Reliability and Validity in Qualitative Research*. Beverly Hills: Sage Publications.
- Kritzer, Herbert M. 1990. "Substance and Method in the Use of Ratio Variables, or the Spurious Nature of Spurious Correlation?" *Journal of Politics* 52(1): 243-254.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.
- MacPhail, Fiona. 1998. "Moving Beyond Statistical Validity in Economics." *Social Indicators Research* 45:119-149.
- Peters, B. Guy. 1991. *The Politics of Taxation: A Comparative Perspective*. Cambridge, MA: Blackwell.
- Ragin, Charles. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64: 1033-1053.
- Schumpeter, Joseph A. [1918] 1954. "The Crisis of the Tax State." In *International Economic Papers*, No. 4, eds. A. Peacock et al. London: Macmillan and Co., Ltd.
- Shah, Anwar and John Whalley. 1991. "The Redistributive Impact of Taxation in Developing Countries." Pp. 166-187 in *Tax Policy in Developing Countries*, eds. J. Khalilzadeh-Shirazi and A. Shah. Washington, DC: The World Bank.
- Steinmo, Sven. 1993. *Taxation and Democracy*. New Haven: Yale University Press.
- Steinmo, Sven and Caroline J. Tolbert. 1998. "Do Institutions Really Matter? Taxation in Industrialized Democracies." *Comparative Political Studies* 31: 165-187.
- Tilly, Charles, ed. 1975. *The Formation of National States in Western Europe*. Princeton, NJ: Princeton University Press.
- \_\_\_\_\_. 1992. *Coercion, Capital, and European States, AD 990-1992*. Malden, MA: Blackwell Publishers.
- Varsano, Ricardo, José Roberto Rodrigues Afonso, Erika Amorim Araujo, Elisa de Paula Pessoa, Julio Cesar Maciel Ramundo, and Napoleão Luiz Costa da Silva. 1998. "Uma Análise da Carga Tributária do Brasil." Rio de Janeiro: IPEA.

- Von Stein, Lorenz. [1885] 1964. "On Taxation." Pp. 28-36 in *Classics in the Theory of Public Finance*, eds. R. Musgrave and A. Peacock. London: Macmillan & Co., Ltd.
- Von Wieser, Friederich. [1924] 1964. "The Theory of Public Economy." Pp. 190-201 in *Classics in the Theory of Public Finance*, eds. R. Musgrave and A. Peacock. London: Macmillan & Co., Ltd.
- Webber, Carolyn and Aaron B. Wildavsky. 1986. *A History of Taxation and Expenditure in the Western World*. New York: Simon and Schuster.
- Weyland, Kurt. 1996. *Democracy without Equity: Failures of Reform in Brazil*. Pittsburgh: University of Pittsburgh Press.
- \_\_\_\_\_. 1998. "From Leviathan to Gulliver? The Decline of the Developmental State in Brazil." *Governance: An International Journal of Policy and Administration* 11: 51-75.
- World Bank. 1988. *World Development Report: Issues in Public Finance*. Washington, DC: Oxford University Press.
- \_\_\_\_\_. 1998. *World Development Indicators* (CD-ROM).