Income taxation and the validity of state capacity indicators

MELISSA ZIEGLER ROGERS
Department of Politics and Policy, Claremont Graduate University, Claremont, USA
E-mail: melissa.rogers@cgu.edu

NICHOLAS WELLER
Department of Political Science and School of International Relations, University of Southern California, Los Angeles, USA
E-mail: nweller@usc.edu

Abstract: State capacity is a key concept for research in public policy and political science. Despite its importance, there is no broadly accepted measure of state capacity in the existing literature, and frequently used measures of capacity have not been examined for their validity. We begin with an explicit definition of state capacity – the state's ability to implement public policy – and connect this definition to a measurable outcome of state capacity – the state's taxation of income. We show that this measure, income taxes as a percentage of total tax revenue, is a useful indicator of state capacity and meets higher standards of measurement validity than other tax-based indicators. We also compare our measure to the most common existing indicators of state capacity to show that income taxation is a better theoretical and statistical measure of states' effectiveness in policy implementation.

Key words: comparative public policy, measurement validity, state capacity, taxation

Introduction

The quality of policy implementation varies widely across and within countries. For scholars in the growing field of comparative public policy, this is a very important and highly relevant topic. Low state capacity can result in outcomes such as lack of access to potable water, poorly developed transportation systems and widespread crime. Even among countries with relatively high-quality policy implementation, such as the US, state capacity has emerged over time and, most argue, within the last century (Skocpol 1982; Carpenter 2001). Understanding the ability of states to implement policy – that is, their state capacity – is thus a central concern of social science.

183

State capacity is widely discussed by scholars in public policy and political science, but there is no consistent definition nor agreed upon measure of the concept. In this paper, we review the prominent definitions and measures of state capacity and demonstrate that an underlying theme in the various definitions is the ability of a state to implement public policy. We adopt a definition of state capacity that focuses on policy implementation and propose to measure the construct by examining the proportion of total tax revenue from income taxes. We make three specific contributions in this paper. First, unlike other scholars, we focus our measure explicitly on income tax collection rather than overall tax collection. Second, we demonstrate the construct validity of our measure. Third, we compare our measure to the most commonly used indicator for state capacity, total tax revenue as a percentage of gross domestic product (GDP), to demonstrate the theoretical and statistical differences between the measures and how ours is a better measure of state capacity for a large sample of cases.

The most important argument of this article is that the choice of state capacity measures depends on the theoretical construct being measured and the ability of a measure to capture this construct empirically. The various indicators of state capacity that we examine in this paper represent different underlying theoretical constructs and measures and, therefore, they are not simply substitutes for each other. Scholars need to think seriously about what they mean by state capacity when choosing a measure for their research. This discussion follows in the tradition of Sartori (1970) and Adcock and Collier (2001) who encourage scrupulous examination of the measurement of key concepts in social science. As with other contested concepts such as democracy (Bollen 1980; Munck and Verkuilen 2001), the appropriate construct and measure must be clearly described and tested for its validity before we can accept the findings that come from quantitative research. The income tax measure that we advocate captures one construct of state capacity that we see as both very common in the literature and not captured by existing measures. This construct and measure is applicable to many fields, especially public policy research.

The article proceeds as follows. In the "What is state capacity?" section, we define state capacity with reference to supporting literature. In the "Our measure" section, we present our income tax measure of capacity and review the most common tax-based measures of state capacity. In the "Construct validity of income tax measures" section, we demonstrate construct validity of our measure by explicating its theoretical justification and "face validity". We also show the empirical validity of the measure by directly comparing it with the most commonly used tax-based measures.

In the "Assessing alternative specifications of taxation measures" section, we address alternatives to the empirical construct of our measure. The last section concludes.

What is state capacity?

A single, clear definition of state capacity is largely absent in policy research. However, many studies that utilise state capacity implicitly adopt a definition of capacity focused upon the state's ability to effectively implement policy. Various definitions meant to break down the concept, such as "top-down" theories of capacity by autonomous or coercive states (Johnson 1982; Haggard 1990; Grindle 1996; Carpenter 2001) and theories of "synergy" between state and society (Migdal 1988; Evans 1995), differ in their views on the drivers of capacity, not on the definition of capacity as effective policy implementation. At their core, definitions of state capacity emphasise the ability of states to enforce policy, even against the preferences of citizens (Mann 1993).

Building upon this literature, we define state capacity as the ability of a state to implement public policy. If a state wants to change the status quo by enacting and enforcing a policy choice and it cannot do so, then it lacks capacity. A state capable of implementing policy, even or especially against societal resistance, is a high capacity state. In this definition, a high capacity state is able to effectively implement a wide range of policies. Importantly, even the most powerful state will not be able to put into place every type of policy or execute all policies perfectly. Our definition does not suggest this. Rather, it focuses on a state's ability to implement policy relative to the ability of other states or to itself over time. A highly capable state will be better able to implement policy than a less capable state.

State capacity in public policy research

Recent public policy research increasingly recognises the importance of examining state capacity to give a full picture of the policy process and policy evaluation. Several trends in public policy analysis highlight the need for state capacity measures, including comparative public policy (particularly in the developing world), sub-national policy research and historical public policy. In all of these contexts, analysts must question the most basic assumptions about technical and political capacity in order to explain variance in outcomes. For example, in a recent prescriptive article on the age of globalisation, Farazmand (2009) emphasised capacity-building as critical to growth and development in the world. Policy studies in Africa (Mead 1996), Eastern Europe (Staronova 2010), Asia (Tsao 2009) and cross-nationally (Kettl 1997) feature capacity or

capacity-building as the most important issue in comparative bureaucracy. Moreover, studies of state and local policy, even in developed nations such as the US, critique the ability of those governments to implement legislated policies (cf Reeves 1982). Despite the importance of capacity measures in public policy research, especially for establishing base levels of capacity in comparative context, little analysis of the validity of these measures has been undertaken.

Our measure

We operationalise state capacity as individual and corporate income taxes divided by total central government tax revenue. To create the income tax measure, we use the average value of total income taxes and total tax revenue for a three-year period spanning the years 2006–2008, which helps to remove minor year-to-year variation. We draw our data on tax collection from the IMF's *Government Finance Statistics* Consolidated Central Government data set.[1] Our measure of state capacity ranges from 0.04 to 0.97, and the numerator includes all individual and corporate income taxes. In Table 1, we present summary statistics, and, in Figure 1, we show a histogram of the distribution of the state capacity measure. The measure ranges from almost no income tax revenue to almost all central government revenue coming from income taxes. Figure 1 suggests that our measure captures tremendous variation in state capacity.[2]

Our measure builds upon Tilly's (1978) assertion that state capacity refers to the power of raising revenue. Additionally, effective taxation, unlike most public policies, can be directly observed and measured. Also, in contrast to other policies such as universal education or national healthcare, the decision of whether to tax citizens is not a matter of debate

---

[1] There are always concerns about the quality of the cross-national economic data. First, there is a considerable amount of missing data, which is likely to be non-random (Hollyer et al. 2011). We expect, but do not investigate in this paper, that data on income taxation is more likely to be missing in countries with low state capacity, the same countries where income taxes are a small proportion of total taxes. We argue that our construct of state capacity is valid across countries and time periods; therefore we do not expect these missing data to affect significantly our results or findings about the validity of our construct and measure. In the denominator of our measure, we use total tax revenue rather than total government revenue. The rationale for this choice is that we believe we understand the theoretical relationship between different forms of tax revenue and state capacity; however, we lack theoretical expectations about the relationship between non-tax revenue and state capacity, so we exclude such sources from our analysis.

[2] All tax data are from the IMF's Government Finance Statistics (IMF 2012). The data used in this paper are available online at melissazieglerrogers.wordpress.com and dornsife.usc.edu/weller. The outliers in Figure 1 are Brunei, Qatar and the US.

Table 1. Summary statistics for income tax measure

	N	Mean (SD)	Minimum	Maximum	25th percentile	75th percentile
Income tax/total tax	119	0.35 (0.18)	0.04	0.97	0.22	0.46
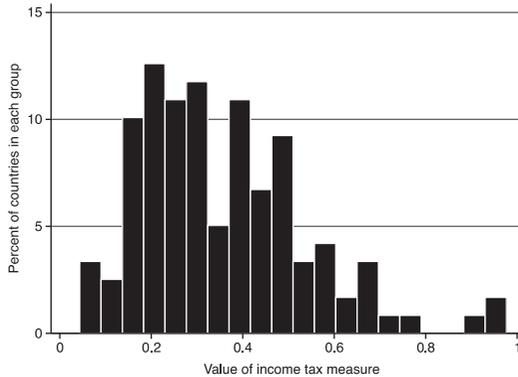
Figure 1 Income tax/total tax data distribution.

for nearly all modern states (Rose 1985). Weyland (1998) succinctly describes these important qualities of taxation:

> Since taxation is a core task of the state, states embark on especially determined action in this area… Thus taxation is one of the "least likely" cases for a decline in state capacity. If such a decline can nonetheless be documented in this area, it is probably even farther reaching in other areas (p. 53).

While this sentiment is common in state capacity literature, taxation data is not widely used as a capacity indicator in public policy research, and the focus on income taxation, rather than total taxation, is mostly missing in the social science literature. Moreover, no studies have demonstrated, quantitatively, that taxation is a good construct for capacity. The existing literature that uses tax collection as a measure of state capacity typically only discusses the face validity of the measure. Face validity helps to show that the measure is a reasonable representation of the construct, but arguments for face validity do not provide evidence that the measure is empirically valid, which is necessary for us to have confidence in the measure. After we briefly review other measures of state capacity and

argue they are not sufficient for our use, we demonstrate construct validity of our measure.

Existing tax-based measures of state capacity

Taxation-based capacity measures examine observable behaviour to capture evidence of successful policy implementation. Our measure follows in the tradition of these taxation-based indicators and the broader fiscal sociology literature that views taxation as a window to state−society relationships. Importantly, the indicators that we review in this section and examine in subsequent sections are distinct, both in theoretical construct and empirical measurement, from our measure and from each other.[3]

The most common tax-based measure of capacity is total tax revenue as a percentage of GDP (Peters 1991; Steinmo 1996; Cheibub 1998). In this formulation, the ability to collect a large proportion of revenue relative to the size of the economy provides evidence that a state is able to extend its authority over economic activity. We agree with the intuition to use actual state behaviour to measure capacity, but we propose an improved measure based on the *properties* of different taxes. The total tax measure includes all types of government taxes, both the easy and difficulty taxes, obfuscating important differences in capacity required for collection. Customs taxation, excise taxation and some forms of sales taxes are achievable by nearly all states. Total taxation as a percentage of GDP, then, is too coarse a measure, both theoretically and empirically, that we can improve upon through a tax properties construct. We focus our

---

[3] We focus on taxation-based measures of capacity in this article, but some scholars use survey-based measures or indexes of capacity-related elements to measure capacity. Some of the more popular surveys are the Business Environment and Enterprise Performance Survey published by the World Bank, the International Country Risk Guide published by the Political Risk Services Group and Transparency International's Corruption Perceptions Index (World Bank 2008; Political Risk Services Groups 2008; Transparency International 2008). In these survey designs, country experts, bureaucrats or business people are asked their opinions about the policy environment in a country. Researchers use data from selected survey responses or they develop a composite indicator of questions across several surveys. We view these as fundamentally different constructs and methodology outside the scope of our argument. However, we also worry about construct validity for these measures. These surveys are typically intended to gauge the business climate or political risk in a country, which is conceptually distinct from the ability of a state to implement public policy. Surveys provide information about the business climate (especially for foreign investment) by collecting opinions on the likelihood of bureaucratic red tape or bribery, or failures in basic public services such as utilities. A state can be very effective in implementing policy that harms the business environment and be said to have strong state capacity. State capacity scholars have recognised a variety of problems with existing surveys and with cumulative indexes (World Bank 2000; Knack 2006).

demonstration of validity on comparing this measure (total taxes/GDP) to our measure (income tax/total tax) in the next section.

Relative political capacity (RPC) is another widely used capacity measure, particularly in international relations research (Organski and Kugler 1980; Arbetman and Kugler 1997). RPC is a measure of actual taxation compared to the amount of tax a country is expected to collect based on the structure of its economy. This measure of capacity is very useful for cross-country and inter-temporal analysis of tax extraction, but it is a different construct than state capacity as policy implementation.

RPC does not incorporate the difficulty of tax administration in its construction. Included in this measure is all revenue, including taxes that are easy to collect. Rather than focusing on how much a state taxes overall, we view the difficulty of tax types as providing valuable information about state capacity. Accordingly, a state that collects a very large amount of revenue from trade and natural resources relative to its economic structure would be high capacity according to RPC.[4] Unless it also extracted significant wealth from income taxes this state would be low on our measure. Lieberman (2004) supports the exclusion of natural resource revenue in measures of state capacity, because it requires "little technical or political capacity to procure it" (p. 98). The existing measures do not make use of the properties of taxes that relate to the difficulty of certain taxes, namely those on income.[5]

We also build upon past capacity measures that utilised income taxation but did not demonstrate construct validity. For example, Chaudhry (1997) sees direct taxation, especially income tax collection, as vital to state building. Levi's (1988) exploration of "quasi-voluntary compliance" focuses primarily on income taxes because the high enforcement costs of incomes taxes make them particularly strong tests of citizen cooperation with states and state force. Neither author analyses whether this construct is an accurate measure of what they intend to study. Our measure of state capacity is clearly in the same family as other measures based on tax collection. However, as we reveal in the theoretical and empirical sections

[4] RPC includes natural resources and exports into its calculation of the structure of the economy but does not distinguish between tax types in its revenue calculation.

[5] The empirical differences in these measures are shown in an online appendix. RPC is shown to have a positive and significant correlation with trade tax revenue/GDP. This makes sense given their construct but not ours. Moreover, RPC is empirically unrelated to several variables we view as important to capacity, especially infrastructure development (Appendix Table A2). Finally, RPC is correlated with several variables we view as unrelated to policy implementation, such as the size of the military in the labour force, the population density and the size of the military budget (Table A3). Thus, we view RPC as capturing a distinct concept from the state capacity as states' effectiveness at policy implementation. Our point of departure from RPC is thus theoretical and empirical.

below, we are proposing a distinct measure of state capacity and, unlike previous studies, we demonstrate construct validity for this measure.

Construct validity of income tax measures

In this section, we demonstrate construct validity for our measure to show it is a valid indicator for state capacity. Showing construct validity is necessary for any good measure; this has been performed for many of the most utilised measures in social science. For instance, Perry (1996) evaluated the construct reliability and validity of measures of public service motivation. Poole and Rosenthal (1985) explicated the validity of their NOMINATE measure of legislator ideology, and Berry et al. (1998) scrutinised measures of citizen and government ideology in the American states.

Shadish et al. (2002) argue there are two fundamental requirements for construct validity. First, researchers must start with a well-explicated construct of interest by specifying exactly what is being measured and how it represents the definition. Second, researchers must be able to determine whether there is a reasonable match between the construct and, their data; they must go beyond arguing that the construct is reasonable and actually show that it is so. We demonstrate the construct validity of state capacity according to these guidelines.

As we elaborated in the "Introduction" section, our definition focuses on the state's ability to implement policy. Our measure – income tax collection – captures a state's broader ability to implement policy and, in this section, we present evidence that income tax revenue is a valid measure. We provide evidence for construct validity by arguing that the measure has face validity and demonstrating that the income tax measure has the predicted pattern of empirical correlations with other outcomes.

Face validity: collecting taxes requires effort and capability

Face validity necessitates that a measure of a given construct is, "on its face", a reasonable measure. In this section, we provide a thorough discussion of income taxation to demonstrate face validity. There are three key points to our argument for face validity. First, we believe that tax collection, in general, is a good starting place for a construct of state capacity, because collected taxes represent the outcome of an implemented state policy. Second, we focus on income taxes as a more specific measure of state capacity, because income taxation is a particularly difficult type of tax collection and therefore provides a more refined measure of the concept. Third, we argue that a state's ability to collect income

taxes will be related to the state's ability to implement other types of policy.

Our indicator of state capacity assumes that states need and want to collect revenue. This is a reasonable assumption, because, without revenue, a state cannot accomplish any tasks it deems necessary. Furthermore, this assumption is consistent with many theories of the state (North 1981; Tilly 1992) and with other measures of state capacity (Levi 1988; Therkildsen 2000). The methods used to acquire fiscal resources, however, are a political choice. These choices, we argue and demonstrate, vary systematically with states' capabilities, making it a useful indicator of state capacity.

Taxation requires substantial effort from states, and income taxation requires the highest level of effort (Atkinson and Stiglitz 1975; Slemrod 1990; Tanzi 1991; Lieberman 2004). An extensive literature in political science and economics has demonstrated the importance of taxation as a crucial linkage between state and societal actors (Musgrave 1969; Levi 1988). States must provide services to citizens in exchange for tax resources, and states must therefore interact with citizens to monitor economic behaviour and to collect and spend tax resources. Taxation of any kind, however, requires a functional state bureaucracy. The types of taxes collected require widely differing levels of effort by the state and its bureaucratic apparatus (Aizenman and Jinjarak 2009). Moreover, Besley and Persson (2009) argue that financial capacity, or taxation, and legal capacity are inseparable. All state capacity, they argue, is dependent on taxation and measure capacity with income taxation. As we argue below, income taxation is a particularly compelling measure of state capacity, because it is lucrative for states but requires substantial capabilities to implement.

Income taxation is an indicator that can capture many diverse aspects of state capacity related to policy implementation. The properties of income tax collection – difficult to administer and enforce – reveal state capacity. As a cross-national measure, it can capture both the national capabilities, or "the extent of resources at its (the central states') disposal for exercising power via its institutions of control", and the weight of the state, or "how the exercise of state power shapes the society it controls" (Soifer 2008, 235–236).

Capacity indicators based on the properties of taxes are also adaptable. Some types of research will require capacity measures other than national income taxes. Public policy research, for example, often examines regional or sub-national (state/province or municipal) differences in policy outcomes that cannot be captured by national-level income tax indicators. For research where income tax measures are not appropriate,

we advocate using other types of taxes that similarly demonstrate effort by the state. For example, at the sub-national level in developing countries, effective property tax collection reveals successful policy implementation (Bird 1992; Rogers submitted). What is crucial for tax-based capacity measures is the theoretical construct that they reveal effort and effective bureaucratic function in the jurisdiction of interest.

We assume not only that governments seek revenue in general, but that they seek tax revenue. In particular, we presume, all else equal, that states would like to collect income tax revenue, because it has the following properties that states find attractive relative to most other taxes: low variance in year-to-year quantity collected, reach in all sectors of the economy and fewer distortions of economic behaviour.

In order for a government to tax effectively, however, three necessary conditions must be met by a state:

1. Economic actors must be known to the state.
2. The state must be able to determine an amount to tax the actors.
3. The state must be able to extract income using the results of conditions 1 and 2. This implies that economic actors voluntarily comply or the state uses force to extract revenue.

These conditions highlight the importance of information, monitoring and use of force for successful taxation. These conditions are agnostic to the type of taxes collected. Even where only small sectors of a state are taxed, as in so-called petro-states, governments must have knowledge of the actors involved in oil extraction, a formula for taxing oil and the ability to enforce compliance with imposed taxes.

Taxation requires meeting the three conditions above, and it will be relatively more or less difficult to meet these conditions depending on the type of tax implemented. One of the easiest forms to collect are trade taxes. All three conditions can be met for trade taxes with relative ease. For a government to collect a tax on international trade, enforcement is largely limited to officials at borders and ports.[6] Since trade must flow through these critical "gates", government interaction with the actors that must use them provides the information required for conditions 1 and 2 to be met. Condition 3, enforcement and collection, is also less demanding than other taxes, because force can be concentrated at these "gates". The ease of collecting these taxes helps to explain why countries have relied on

---

[6] Of course, the existence and maintenance of borders and ports comes before this. Part of the motivation for governments to assist, develop or control these infrastructures comes from their desire to tax them.

them for considerable amounts of revenue, both historically and currently in many less-developed countries.

Trade taxes may be "easy", but they impose considerable economic costs on a society and have limited revenue potential (Tanzi and Zee 2004). Sales taxes can have high revenue potential, are moderately distortionary and are moderately difficult to collect (Lieberman 2004; Bird and Zolt 2005). Importantly, income tax systems, especially once automated, are the steadiest and most lucrative form of tax revenue. All else equal, most governments would prefer to tax from the widest base possible and to do so in the least distortionary manner.

Of course, all else is not equal, political concerns weigh heavily into decisions about taxation. Nevertheless, if feasible, most governments would want to garner more from "efficient" taxes. The taxes with the widest, most stable bases are those on income and property. These are effectively applied in developed nations; once these governments were able to tax activities that are more difficult to monitor, such as income and goods and services, trade taxes became less important for government revenue. This suggests that states will prefer more difficult, information-intensive taxes if they are feasible to enforce. These taxes, however, require considerable amounts of information to meet the three necessary conditions for taxation.

Empirical demonstration of construct validity

Measures must not only have theoretical validity but also empirical grounding. In this section, we demonstrate empirically that our proposed measure of state capacity is a valid indicator of the underlying construct we identify. We take several approaches. First, we show that our measure of state capacity correlates positively with other observable indicators of high capacity states. We focus on correlations with two types of indicators – measures of infrastructure quality and measures of administrative quality. Both should be positively correlated with a state's ability to implement policies, which is the core of our theoretical construct. Second, we demonstrate that our measure does not correlate with attributes of a state that should be unrelated to our construct of state capacity.

We first discuss how our measure correlates positively with other indicators of a capable state, often called convergent validity (Trochim and Donnelly 2007). We focus, in particular, on activities that are related to a state's basic infrastructure functions. We expect the quality of physical infrastructure to correlate with state capacity, because state actors are usually central to the financing and construction of large infrastructure projects. Moreover, infrastructure in water and sanitation is typically the purview of states. We expect each measure to vary positively

with state capacity, because basic infrastructure improvements require a government that can implement public policy.

In addition to the infrastructure indicators, we examine the correlation between our measure and the regulatory quality variable from the World Governance Indicators (Kaufmann et al. 2009). This variable is defined as "perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development". The focus of this measure on policy implementation is appropriate given our construct for state capacity. We also examine our measure's correlation with the government effectiveness variable from the same data set. This variable attempts to capture "perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies". Both of these measures capture legal and financial infrastructure, which Besley and Persson (2009) view as inseparable from state capacity. Although these measures of institutional quality are imperfect, they are commonly used in research and, importantly, we did not collect the data or design the indicator, thus it is independent validation of our measure.

We turn now to correlations between our measure of capacity and other variables that should be positively correlated with a state's ability to implement public policy. In assessing a measure's validity, it is important to understand how it correlates with other relevant variables, and it is also important to consider how alternative measures correlate with those other data. To that end, we also examine how total tax/GDP relates to these other correlates of state capacity.

For all of our variables, we compute the average value for a three-year period spanning the years 2006–2008, which helps to remove minor year-to-year variation. If a particular country has data for only one or two of the three years, we still retain them in the comparisons. In Table 2, we report the correlations between the two primary measures of state capacity and the infrastructure and the World Governance Indicators measures.

The data reveal what we expect – consistent correlations between the income tax measure and measures of infrastructure quality. This is empirical evidence that income tax collection is related to a government's ability to actually implement public policy. Although the magnitude of the correlations varies across the different variables, the correlations are uniformly positive and significant, as we would expect if our state capacity measure captures the ability of a state to implement policy. Just as importantly, the total tax measure features consistently lower correlations with these different empirical measures, and at times the correlations are either insignificant or in the opposite direction. For instance, in Table 2,

Table 2. State capacity indicators and state attributes

	Improved water source (% population)	Improved sanitation (% population)	Total km of roads	Per cent of total roads that are paved	World Governance Indicators, government effectiveness	World Governance Indicators, regulatory quality
Income tax/total tax	0.28 (0.01, 114)	0.33 (0.01, 112)	0.33 (0.01, 78)	0.12 (0.33, 65)	0.48 (0.00, 118)	0.4600 (0.00, 117)
Total tax/GDP	0.11 (0.25, 113)	0.13 (0.17, 111)	−0.22 (0.05, 78)	0.26 (0.04, 65)	0.22 (0.02, 117)	0.1676 (0.07, 117)

In each cell, we report the pairwise correlation coefficient and in parentheses the level of significance of the correlation and the number of observations.

there is a negative correlation between the total tax measure and the total kilometres of roads in a country, which indicates that total taxation does not correlate with road infrastructure, an indicator for state reach in policy making and evidence of one type of policy delivery. Total tax/GDP is also unrelated to other measures of health policy implementation, including improving sanitation and access to potable water. The total tax measure of state capacity fails to show a consistent correlation with the other outcomes that a high capacity state will likely achieve.[7] Therefore, if one accepts that a measure of state capacity should correlate with other attributes of an effective state, these results suggest that the income tax/total tax measure is superior to the most common alternative measure of state capacity. It may be that total taxation is a good measure for another concept of state capacity, but it does not seem that the total tax measure is as good a measure of the construct we identify for state capacity.

Turning now to Table 3, we examine whether the measures of state capacity correlate with variables that should *not* be correlated with capacity. The crucial task in this empirical exercise is to choose variables for which a correlation would suggest that our measure of state capacity actually captures something that is unrelated to, but correlated with, state capacity (Trochim and Donnelly 2007). In this vein, we examine two measures of population (total population and population density) and two measures of military strength (size of the armed forces and military spending) from the World Development Indicators (World Bank 2012).

The results in Table 3 show that the income tax measure correlates with neither the population measures nor the military measures, but the total tax measure does correlate with total population and total military spending. The absence of a correlation between our income tax measure and the population measures is important, because tax-based measures could correlate with either of these measures for reasons unrelated to a state's actual capacity. It is plausibly more difficult to collect taxes when populations are large and density is low, which places a greater administrative burden on the tax collection bureaucracy. Therefore, the absence of a correlation with the population variables demonstrates that it is not simply more densely populated states that are able to tax income at a higher level. A correlation between density and income taxation would call into question whether or not our measure captured the state's actual capacity. If the correlation existed, it would suggest that greater income taxation is a function of demographic variables rather than the actions of

---

[7] We only show the results comparing our measure to the most commonly used measure, total tax/GDP, for reasons of space and clarity. We have included the results for these tests on RPC, income/GDP and VAT in the online appendix in Tables A2 and A3.

Table 3. State capacity indicators and theoretically unrelated attributes

	Total population	Mean population density	Armed forces personnel (% of labour force)	Military spending (% of government spending)
Income tax/total tax	0.09 (0.33, 119)	−0.13 (0.17, 119)	0.0071 (0.94, 113)	0.0180 (0.86, 101)
Total tax/GDP	−0.1961 (0.03, 118)	0.0326 (0.73, 118)	0.0634 (0.51, 112)	0.2079 (0.03, 109)

In each cell, we report the pairwise correlation coefficient and in parentheses the level of significance of the correlation and the number of observations.

the state. It is illustrative, therefore, that total tax/GDP is significantly correlated with one of the population measures. This suggests that these measures may be tapping into attributes of a state that reflect geographic or demographic characteristics more than state action.[8]

The income tax measure is also not statistically related to the measures of military size and spending. This is consistent with our measure's focus on policy implementation, which is not inherently related to military capability. Total tax/GDP is correlated with military spending, suggesting that these variables are either tapping into attributes of a state that are not related to state capacity or that their construct is different – they want to capture an alternative definition of capacity (perhaps capacity for fighting war) rather than public policy implementation.

These results show that income taxes as a proportion of total government taxation do not correlate with measures that should be unrelated to state capacity. This is an important, if somewhat unusual, way to demonstrate the value of a proposed measure. If we had found that income taxation had a significant correlation, with indicators for which we did not expect a correlation then it would appear that some other, unmeasured variable determined both income taxation and land usage and military attributes of a state. The empirical results provide consistent evidence that our measure captures state capacity, and that the total tax measure does not as effectively capture the policy implementation aspect of state capacity.

The last step we take in this paper is to compare directly our proposed measure of capacity to the measure based on total government tax revenue divided by GDP. The correlation between the two measures is only 0.07 for the three-year average of 2006–2008, which means that the two measures are not only theoretically but also empirically distinct. The very low correlation between the two measures makes it imperative to use the measure that is most appropriate for a given construct.

To investigate the validity of the two measures, we examine the countries where the two measures are most divergent from each other. To do so, we determined the countries that are above the 75th percentile on one measure and below the 25th percentile on the other measure. This creates two groups of countries – one group that is high capacity on the income tax measure but low capacity on the total tax measure and one group that is low capacity on the income tax measure but high capacity on the total tax measure. Although the 25th and 75th percentile cutoffs are arbitrary, the goal is to identify the countries for which the two rankings are widely different. The countries

---

[8] Of course, we are not denying that demographic and geographic factors influence states' ability to implement policy, but the concept of state capacity should be analytically distinct from these variables.

ranked as high capacity on the income tax measure but low on the total measure include: Bhutan, Canada, Iran, India, Japan, Kazakhstan, Singapore, Spain and the US. The countries ranked as high capacity on the total tax measure, but low on the income tax measure include: Argentina, Belarus, Benin, Cuba, Estonia, Macao, Madagascar and Lesotho.

The question now becomes, which measure seems more appropriate for each group of countries? It is difficult to answer that definitively, but we question the validity of a measure that places the US, Canada and Japan as among the world's least capable states. At the same time, the total tax measure ranks Lesotho and Madagascar as among the most capable states. These results strike us as, on face, problematic for the total tax measure. In addition, these groupings highlight the theoretical challenges of the highly aggregated total tax measure. Lesotho, for example, is highly dependent on trade taxation (over 40 per cent of total taxation) to meet its needs. This characteristic of a weak state makes Lesotho appear capable on the total tax measure. Lesotho, moreover, performs very poorly on many indicators of high quality policy implementation, especially health outcomes and infrastructure development, relative to their wealth and their neighbours (CIA World Factbook 2012). Argentina also appears misplaced in these rankings. While not a failing state, it is widely regarded as having low taxation capacity relative to the world, its neighbours and its level of development (Bergman 2003).

Assessing alternative specifications of taxation measures

Our measure of income tax/total taxes is only one possible way to utilise the income tax construct. In this section, we specify different ways we could have measured the numerator as well as alternatives to total tax as the denominator. We also discuss why we did not use the value added tax (VAT) to capture difficult taxation, rather than income taxes. We emphasise the theoretical reasoning behind our choices with references to empirical findings in the online appendix.

Why not individual income tax as the numerator?

Total income tax values include collection of individual income taxes and corporate income taxes. These two taxes have different incidence – the former is collected from all eligible citizens and the latter from businesses. In terms of state reach and administrative difficulty, the individual income tax may be the most challenging tax a state collects. Scholars could argue on this basis that the corporate tax dilutes the income tax variable because it is easier.

If there are states that collect substantial corporate income taxes but minimal personal income taxes, then, by combining the two elements into

the numerator, we might unintentionally claim that states that collect taxes from businesses (easier) are as equally capable as states that collect taxes from individuals (harder). Overall, the relationship between the ratio of total corporate income taxes to GDP and the ratio of total individual income taxes to GDP show a correlation of only 0.07.[9] To further examine the different types of income taxes, we again look at countries that are highly ranked by the corporate income tax measure but ranked low on the income tax measure. The countries that are above the 75th percentile on corporate income tax collection and below the 25th percentile on individual income tax collection are: Cuba, Brunei, Kazakhstan, Trinidad and Tobago, and Uruguay. If the concern about tax incident is correct, then our measure using total income taxes may overstate the capability of these states. But since the list above includes states that are both high and low capacity on our total tax measure, it is unclear if the total tax measure is consistently wrong for these states or if there are few states ranked high on the corporate measure but low on the individual measure.[10] This gives us confidence that combining the two types of income taxes does not weaken our measure.

Aside from the empirical evidence, it is not obvious that collecting corporate income taxes is easy in theory. The corporate income tax involves fewer actors, but it is arguably more challenging politically and administratively due to sophisticated corporate avoidance or evasion methods. The business sector is better organised to counter income tax claims and better able to use the tax code to their advantage (Slemrod 2007). It is also likely important for state capacity that the state is able to extract from these powerful actors. For both empirical and theoretical reasons, we believe the choice to combine corporate and income taxes in the numerator is justified.

Why not GDP as the denominator?

Income taxation, as the numerator in a capacity ratio, is the important innovation for a measure of state capacity that we stress in our research. The choice of denominator for income tax revenue, however, is also consequential for the measure of state capacity. We argue that income

---

[9] We use GDP as the denominator rather than total income taxes, because clearly corporate and individual income taxes are substitutes for each other when total income taxes are the denominator. This denominator (GDP) allows us to identify if the two are related relative to the overall size of the economy.

[10] Brunei, for example, does not collect an individual income tax, capital gains taxes, trade taxes, payroll taxes, sales taxes or manufacturing taxes. Brunei's is thus an outlier, because it is a tax haven that only collects corporate income taxes and selected other minor taxes (Brunei Economic Development Board).

taxes/total tax collection is the best theoretical measure to capture our understanding of state capacity, therefore challenging both the numerator and denominator of the most popular measures of capacity.

Other tax-based measures of state capacity use either total taxes (or revenue) as the numerator in a ratio but typically use a measure of economic size (i.e. GDP) as the denominator. We do not adopt this approach in our measure, because it conflates normative and positive theories of state capacity. Measures that utilise an economic outcome (such as GDP) in the denominator are essentially measuring how much of the economy that state chooses to collect in tax revenue. This is true whether the numerator is total taxes, income taxes or some other type of tax-based measure. In this view, capacity is a function of the amount of the total economy that the state extracts. However, this measure is subtly normative; that is, different views of the appropriate relationship between the state and society have different implications about the ideal size of the relationship between the government's tax collection and the economy.

We prefer a measure that does not make claims that a larger government is also a more capable government. Rather, our measure addresses the following question: conditional on the size of the government sector, how does the government choose to pay for its activities? We view the structure of revenue as consequential and income taxes specifically indicative of state capacity.[11] The total tax revenue denominator emphasises the tradeoffs that states face in balancing their revenue needs with the states' means to collect them. Nonetheless, this denominator has drawbacks, namely that it suffers when comparing states with very different incidence of total tax revenue, because even modest income taxation will appear significant when divided by very modest total tax revenue. For comparisons across very different levels of total taxation, therefore, the inclusion of GDP as a control variable or thoughtful research design groupings can help to calibrate comparisons across levels of development (Lieberman 2004).[12]

---

[11] The choice between denominators is primarily theoretical but does not appear too consequential, empirically. Income/total tax and income/GDP are highly correlated (online appendix, Table A1) and yield similar results in our construct validity tests, even if income/total tax is a better predictor of high capacity policy implementation (online appendix, Table A2).

[12] The drawbacks of the GDP denominator for measuring capacity are well explained by Lieberman (2004): "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" (pp. 106–107).

Why not the VAT?

Income taxation is widely believed to be the most difficult tax to enforce because of its administrative complexity and the reach it requires into society. Those familiar with complex VATs, common across the globe, might reasonably suggest that effective collection of the VAT is also an indicator of high state capacity. We argue that VAT collection is a weaker indicator of state capacity in theory – the VAT is easier to collect than income tax (Slemrod 1990; Lieberman 2004; Bird and Zolt 2005; Bird and Gendron 2006; Di John 2006). Lieberman (2004), for example, argues that domestic consumption taxes, like the VAT, "require extensive coverage, but monitoring/enforcement less than income taxes" (p. 99). Bird and Zolt (2005) specifically argue against using the income tax as a revenue-generating tax in the developing world because of developing countries' limited capacity to collect it in comparison to the easier VAT (pp. 1665–1671).[13]

The specification of these measures is a consequential theoretical and empirical choice. We have offered evidence that our measure has validity for the construct of state capacity as public policy implementation. For researchers with other constructs in mind, alternatives to our measure should be considered. Most importantly, researchers should evaluate the existing measures of capacity to match both their theoretical and empirical connection with the construct of capacity that they have in mind.

Conclusion

In this article we presented a definition of state capacity that focuses on the ability of a state to implement public policy. This is at the core of most prior characterisations of state capacity, but this definition has not been plainly explicated and linked to a measure tested for construct validity. We demonstrated in this paper that measuring our concept of state capacity as the proportion of total tax revenue from income tax collection has a solid theoretical and empirical foundation. Income tax collection is a good measure for our construct of state capacity, because income taxes represent the

---

[13] The VAT, as a measure of state capacity, also presents empirical concerns. VAT collection is correlated with trade taxation/total taxation in the 2006–2008 period (online appendix, Figure Table A1). In the current age of globalisation, states collect significant revenue from trade taxes because they must for reasons of low capacity, not as optimal tax policy (Aizenman and Jinjarak 2009). Importantly, effective VAT collection is not a sign of a weak state. Rather, the mix of taxes is indicative of capacity. If a state taxes VAT and income at high levels, it is a high capacity state. However, if you tax VAT at a high level but are unable to tax income effectively, this is a sign of a low capacity state.

outcome of particularly difficult policy implementation. We demonstrate construct validity for our proposed measure through face validity, convergent validity and discriminant validity. Importantly, we show that our measure is a better representation, theoretically and statistically, than the dominant measures of capacity focused upon total tax collection. The variety of ways that we demonstrate construct validity gives us confidence that our measure indeed is appropriate for our construct of state capacity.

Foundational concepts in social science research demand careful thought and testing. Our construct and measure of state capacity is one potential construct of the concept that we believe is widely accepted in public policy research and many other fields. As we have demonstrated theoretically and empirically, the choice of state capacity indicators is important, because there are considerable differences between the different indicators. Researchers should thus clarify the concept they intend to capture and choose the measure that best captures that concept. We have shown that, if scholars want to study the ability of a state to implement policy, they would be well served to utilise a measure based on the state's ability to collect income taxes.

Acknowledgements

Supplementary materials

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0143814X1300024X

References

Adcock R. and Collier R. (2001) Measurement Validity: A Shared Standard for Quantitative and Qualitative Research. *American Political Science Review* 95(3): 529–546.

Aizenman J. and Jinjarak Y. (2009) Globalisation and Developing Countries: A Shrinking Tax Base? *Journal of Development Studies* 45(5): 653–671.

Arbetman M. and Kugler J. (1997) *Political Capacity and Economic Behavior*. Boulder, CO: Westview Press.

Atkinson A. B. and Stiglitz J. (1975) The Design of Tax Structure: Direct Versus Indirect Taxation. *Journal of Public Economics* 6(1–2): 55–75.

Bergman M. S. (2003) Tax Reforms and Tax Compliance: The Divergent Paths of Chile and Argentina. *Journal of Latin American Studies* 35(3): 593–624.

Berry W. D., Ringquist E. J., Fording R. C. and Hanson R. L. (1998) Measuring Citizen and Government Ideology in the American States, 1960–93. *American Journal of Political Science* 42: 327–348.

Besley T. and Persson T. (2009) The Origins of State Capacity: Property Rights, Taxation, and Politics. *American Economic Review* 99(4): 1218–1244.

Bird R. (1992) Tax Reform in Latin America: A Review of Some Recent Experiences. *Latin American Research Review* 27(1): 7–36.

Bird R. and Zolt E. (2005) Redistribution Via Taxation: The Limited Role of the Personal Income Tax in Developing Countries. *UCLA Law Review* 52(6): 1627–1696.

Bird R. M. and Gendron P. (2006) Is VAT the Best Way to Impose a General Consumption Tax in Developing Countries? *Bulletin for International Taxation* 60(7): 287–296.

Bollen K. (1980) Issues in the Comparative Measurement of Political Democracy. *American Sociological Review* 45: 370–390.

Brunei Economic Development Board. http://www.bedb.com.bn/doing_guides_taxation.html (accessed 17 October 2012).

Carpenter D. (2001) *The Forging of Bureaucratic Autonomy: Reputations, Networks and Policy Innovation in Executive Agencies, 1862–1928*. Princeton, NJ: Princeton University Press.

Central Intelligence Agency. *WorldFactbook*. Online access 17, October 2012.

Chaudhry K. A. (1997) *The Price of Wealth: Economies and Institutions in the Middle East*. Ithaca, NY: Cornell University Press.

Cheibub J. A. (1998) Political Regimes and the Extractive Capacity of Governments: Taxation in Democracies and Dictatorships. *World Politics* 50: 349–376.

Di John J. (2006). The political economy of taxation and tax reform in developing countries (No. 2006/74). Research Paper, UNU-WIDER, United Nations University.

Evans P. B. (1995) *Embedded Autonomy: States and Industrial Transformation* (p. 12). Princeton, NJ: Princeton University Press.

Farazmand A. (2009) Building Administrative Capacity for the Age of Rapid Globalization: A Modest Prescription for the Twenty-First Century. *Public Administration Review* 69(6): 1007–1020.

Grindle M. S. (1996) *Challenging the State: Crisis and Innovation in Latin America and Africa*. Cambridge: Cambridge University Press.

Haggard S. (1990) *Pathways from the Periphery: The Politics of Growth in Newly Industrializing Countries*. Ithaca, NY: Cornell University Press.

Hollyer J., Rosendorff B. P. and Vreeland J. (2011) Democracy and Transparency. *The Journal of Politics* 73(4): 1191–1205.

International Monetary Fund (2012) *Government Finance Statistics*, CD Access.

Johnson C. (1982) *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925–1975*. Stanford, CA: Stanford University Press.

Kaufmann D., Kraay A. and Mastruzzi M. (2009) Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996–2008. World Bank Policy Research Working Paper (4978).

Kettl D. (1997) Global Revolution in Public Management: Driving Themes, Missing Links. *Journal of Policy Analysis and Management* 16(3): 446–462.

Knack S. (2006) Measuring Corruption in Eastern Europe and Central Asia: A Critique of Cross-Country Indicators. World Bank Policy Research Working Paper (3968).

Levi M. (1988) *Of Rule and Revenue*. Berkeley, CA: University of California Press.

Lieberman E. S. (2004) Taxation Data as Indicators of State-Society Relations: Possibilities and Pitfalls in Cross-National Research. *Studies in Comparative International Development* 36(4): 89–115.

Mann M. (1993) *The Sources of Social Power*. Cambridge: Cambridge University Press.

Mead T. (1996) Barriers to Local Government Capacity in Nigeria. *American Review of Public Administration* 26(2): 159–173.

Migdal J. (1988) *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton, NJ: Princeton University Press.

Munck G. and Verkuilen J. (2001) Conceptualizing and Measuring Democracy. *Comparative Political Studies* 35(1): 5–34.

Musgrave R. A. (1969) Cost-Benefit Analysis and the Theory of Public Finance. *Journal of Economic Literature* 7(3): 797–806.

North D. (1981) *Structure and Change in Economic History*. New York, NJ: W.W. Norton and Company.

Organski A. F. K. and Kugler J. (1980) *The War Ledger*. Chicago, IL: University of Chicago Press.

Perry J. L. (1996) Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity. *Journal of Public Administration Research and Theory* 6(1): 5–22.

Peters B. G. (1991) *The Politics of Taxation: A Comparative Perspective*. Cambridge, MA: Blackwell Press.

Political Risk Services Group (2008) International Country Risk Guide.

Poole K. and Rosenthal H. (1985) A Spatial Model for Legislative Voting. *American Journal of Political Science* 29(2): 357–384.

Reeves M. M. (1982) Look Again at State Capacity: The Old Gray Mare Ain't What She Used to Be. *The American Review of Public Administration* 16(1): 74–89.

Rogers M. (Submitted). Taxing with Dictators and Democrats: Regime Effects, Transfers, and Revenue in Argentina. Under review. *Journal of Politics in Latin America*.

Rose R. (1985) Maximizing Taxation While Minimizing Political Costs. *Journal of Public Policy* 5(3): 289–320.

Sartori G. (1970) Concept Misinformation in Comparative Politics. *American Political Science Review* LXIV (4): 1033–1053.

Shadish W., Cook T. D. and Campbell D. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York, NY: Houghton Mifflin Publishing.

Skocpol T. (1982) State Capacity and Economic Intervention in the Early New Deal. *Political Science Quarterly* 97(2): 255–278.

Slemrod J. (1990) Optimal Taxation and Optimal Tax Systems. *The Journal of Economic Perspectives* 4(1): 157–178.

—— (2007) Cheating Ourselves: The Economics of Tax Evasion. *The Journal of Economic Perspectives* 21(1): 25–48.

Soifer H. (2008) State Infrastructural Power: Approaches to Conceptualization and Measurement. *Studies in Comparative International Development* (43): 231–251.

Staronova K. (2010) Regulatory Impact Assessment: Formal Institutionalization and Practice. *Journal of Public Policy* 30(1): 117–136.

Steinmo S. (1996) *Taxation and Democracy: Swedish, British and American Approaches to Financing the Modern State*. New Haven, CT: Yale University Press.

Tanzi V (1991) Tax Reform in Economies in Transition: A Brief Introduction to the Main Issues. IMF Working Paper (91/23).

Tanzi V. and Zee H. (2004) *Tax Policy for Developing Countries* (Vol. 27). Washington, DC: International Monetary Fund.

Therkildsen O. (2000) Comparative Studies of Relevance to Resource Mobilization and Democratization in African Countries: A Critical Review. Paper presented at the conference The State Under Pressure, The Norwegian Association for Development Research, Bergen, 5–6 October.

Tilly C. (1978) *From Mobilization to Revolution*. Reading, MA: Addison-Wesley.

—— (1992) *Coercion, Capital and European States, AD 990–1990*. Cambridge, MA: Blackwell Press.

Transparency International. (2008) Corruption Perceptions Index.

Trochim W. and Donnelly J. (2007) *Research Methods Knowledge Base*. Cincinnati, OH: Atomic Dog Publishing.

Tsao K. K. (2009) Building Administrative Capacity: Lessons from China. *Public Administrative Review* 69(6): 1021–1024.

Weyland K. (1998) From Leviathan to Gulliver? The Decline of the Developmental State in Brazil. *Governance* 11(1): 51–75.

World Bank. (2000) *Anticorruption in Transition: A Contribution to the Policy Debate*. Geneva: World Bank.

—— (2008) *EBRD-World Bank Business Environment and Enterprise Performance Survey(BEEPS)*. Geneva: World Bank.

—— (2012) *World Development Indicators*. Geneva: World Bank.