Editor's Note

The Journal of Politics was very pleased to publish the exchange on "growth and governance" in our last issue. This is an important debate and each of these essays contributed to it. Some have inquired why there were four essays that included two responses from Kaufmann and his colleagues. Originally, I planned only to have 3 essays: the original article by Kurtz and Schrank, a response by Kaufmann and his colleagues, and then a follow-up by Kurtz and Schrank. But Kaufmann, Kraay, and Mastruzzi wanted a chance to respond, believing that it was only fair to do so. I agreed to their request, especially since Kurtz and Schrank were open to it. Kurtz and Schrank demonstrated their professionalism by going along with this plan. They wanted to see the debate advanced and felt that further attention to the topic would accomplish that. I am raising this matter only as a way of saying that having a 4-part exchange is not any statement by JOP or me about the merits of the positions of these respective scholars. The fact that Kaufmann and colleagues wanted another chance to respond strongly suggests that the critiques offered by Kurtz and Schrank were serious and warranted attention. The contributions of all these scholars are a statement about this debate's importance. JOP was pleased to be part of it.

John Geer

The Journal of Politics, Vol. 69, No. 3, August 2007, p. 595

ISSN 0022-3816

Growth and Governance: Models, Measures, and Mechanisms

Marcus J. Kurtz The Ohio State University
Andrew Schrank University of New Mexico

The regnant scholarly consensus linking good governance—the quality of public administration—to economic development has undergone surprisingly little empirical scrutiny. We examine the relationship by asking two questions: How confident are we in our cross-national measures of good governance? How solid are the empirical foundations of the growth-governance causal linkage? Our results suggest that the dominant measures of governance are problematic, suffering from perceptual biases, adverse selection in sampling, and conceptual conflation with economic policy choices. Within the limits of somewhat problematic measures, the evidence suggests that there is far more reason to believe that growth and development spur improvements in governance than vice versa. The policy implications are profound, for international organizations and governments are beginning to condition developmental aid on problematic measures of administrative performance.

Most analysts agree that political corruption and malgovernance are among the principal barriers to economic development and social betterment in the Third World (see, e.g., Castañeda 2003; Wolf 2005). Conversely, the belief that good governance—the quality of public administration—promotes growth and development is all but entirely uncontroversial (Kaufmann 2005). It forms a framing assumption in a host of academic analyses, a core piece of advice provided by the international financial institutions, and the rationale for new conditions imposed upon recipients of bilateral and multilateral aid (*Economist* 2005; Hopkin 2002; Radelet 2002, 2003; Seligson 2002).[1] In fact, the contemporary paeans to public sector probity are so pervasive as to imply that the link between growth and governance is an article of faith or a starting point for analysis rather than a hypothesis subject to falsification.

Nevertheless, the relationship between the quality of administration and the level or rate of economic development has received little direct scrutiny. Neither a theoretically nor an empirically convincing case for the *beneficial* effects of corruption or malgovernance has been advanced, of course, but the developmental costs are not nearly so clear as the conventional wisdom implies.[2] In fact, we will argue below that we lack genuine consensus as to what malgovernance really *is*; we are further still from cross-nationally valid measures thereof; and we are therefore decidedly premature in assigning causal priority to governance and not vice versa.

What *is* well known is that exceptionally high levels of economic development are associated with what are commonly seen as competent public sectors. We plan to examine the causal status of this correlation. Does good governance actually cause growth? Does economic growth itself promote better governance? Or, are the two phenomena simply the independent products of an underlying, but unmeasured, omitted variable? Do the available measures—i.e., the ones that are in widespread use—give us the tools to directly answer these (and other) questions in a cross-nationally valid way? The answers have important

[1]We want to be very clear at the outset that the phrase "good governance" describes the probity of public administration and implies no particular policy regime or level of public goods.

[2]A small literature suggests that political corruption can facilitate development (see, e.g., Leff 1964 and at least conditionally Huntington 1968, Lui 1985, and Rashid 1981). This is most assuredly *not* our claim.

The Journal of Politics, Vol. 69, No. 2, May 2007, pp. 538–554

 ISSN 0022-3816

theoretical and policy implications, for if economic development or underlying sociopolitical variables which are logically prior to economic development are principally responsible for historical improvements in the quality of governance, and not vice versa, then we should expect institutional reforms that are not accompanied by substantial complementary transformations of society or the political economy to be at most of marginal impact in the quest for good government *and* economic development.

In this paper we advance the simple but novel claim that the relationship between governance and growth rests on far weaker empirical foundations than is customarily claimed. Indeed, we contend that the opposite hypothesis—that is, that economic development drives political modernization—may have more empirical support than the current conventional wisdom implies.

Our potentially controversial claim is based on two types of evidence. First, we examine the best existing measures of the quality of political institutions—the governance indicators recently developed under the auspices of the World Bank (Kaufmann, Kraay, and Mastruzzi [KKM] 2003, 2005). We show that these—and indeed most—indicators that include perception-based measurements of the probity and efficacy of public institutions are quite colored by recent economic performance (see also Seligson 2006, 385), riddled with problems of adverse selection, and feature deeply entrenched biases both for and against various public *policy alternatives* that are logically distinct from the question of public sector effectiveness per se. The consequences are profound, as apparent links between governance and growth are thus more to likely to be artifacts of measurement than reflections of underlying causal dynamics. The second part of the paper seeks to directly evaluate this possibility through a careful examination of the question of causal order and the predictive power of these substantially perception-based measures. The results ratify our concern: Antecedent economic conditions are strong predictors of perceptions of the quality of public institutions, but the *ex ante* measure of governance shows little capacity to predict *subsequent* patterns of economic performance.

What are the ultimate implications? And what should be done? In the first place, we need to be more careful about how we conceptualize governance—to avoid dependence on nonneutral sources and the corresponding tendency to embed policy preferences within the concept. Second, we must build historically valid indicators that allow us to evaluate the growth-governance linkage over longer spans of time. And, finally, we counsel the avoidance of research programs that put the cart in front of the proverbial horse by taking as their starting points the *assumption* that improvements in the institutional rules of government will drive broader socioeconomic development.[3] Instead, we believe that intellectual energy would be better spent in an effort to discern whether in fact it is economic development that drives improvements in governance or allows institutional changes to have practical effect, or, alternatively, if there are unobserved causal factors that select countries into high-growth/good governance or low-growth/malgovernance equilibria (see, e.g., Caselli and Morelli 2003; Mauro 2004). Extant research on malgovernance and corruption has focused quite narrowly on the question of *institutional* context, examining, for example, whether democratic politics, federalism, transformations of administrative structure, or the incentives embedded in electoral or legislative institutions can drive improvements in probity.[4] But if we are right, none of these may be as fruitful avenues of exploration as they seemed at first.[5] Instead, the efficacy of such institutional reforms may be conditional on transformations of underlying economic and social structures that themselves determine the degree to which governments can be held to the goals embodied in such reforms or whether they are yet another in a long series of dead letters. What these structures are and how their effects are manifested is precisely where we think the greatest marginal returns to future investigation lie.

The Theoretical Terrain

The principal empirical research puzzle can be stated simply enough: Does growth underwrite good governance? Does bureaucratic probity promote growth? Or is their strong apparent linkage related to unex-

[3]This does not mean that efforts to combat corruption are not useful in and of themselves. We merely claim that they are likely to be more effective in the context of changes to underlying socioeconomic characteristics that are themselves likely to reduce malgovernance.

[4]See, for example, Geddes (1994) on legislative politics; Weingast (1995) on federalism; and Rose-Ackerman (1999) and Gerring and Thacker (2004) on the structure of political institutions.

[5]Worse, yet, there are potentially pernicious practical consequences. By tying development aid to improvements in governance, the international donor community could well aggravate poverty and inequality. International aid might be directed away from precisely those states that need it most and whose political institutions will in all likelihood resist effective reform without aid.

plored exogenous factors?[6] While the questions are easy to ask, they are hard to answer. To begin with, while growth can be measured in a fairly straightforward fashion, good governance is much more problematic. And to the extent that current explanations suggest that probity promotes *long-term* economic development, we are further constrained to examine historical indicators of good governance. Even where good historical data might be available, evaluating the direction of causality (from growth to governance or the reverse, and in what proportion) relies on our ability to find appropriate instruments that are correlated with, for example, governance, but unrelated to development. This search has proven to be difficult indeed, as nearly all the factors that are related to growth are also typically correlated with measures of governance (Durlauf, Johnson, and Temple 2005; Rodrik 2005).[7]

A substantial and growing body of literature nonetheless holds that governance is more cause than consequence of growth (see, e.g., Kaufmann 2005). After all, Reynolds identified "administrative competence" as the "single most important explanatory variable" (1983, 976) in his magisterial survey of development outcomes in the Third World. Economists and political scientists self-consciously embraced—and *quoted*—his conclusions (see, e.g., Brautigam 1992, 16; Jomo 2000, 345; Riedel 1988, 37; Stern 1989, 614). And they eventually discovered, developed, and deployed cross-national indicators designed to put the growth-governance relationship to the test. Thus, Mauro finds that investment and growth are related to indicators of "bureaucratic efficiency" developed by a private vendor—and portrays the high market price of the indicators as a testimonial to their "accuracy and relevance" (1995, 684; see also Chong and Calderón 2000). Gupta, Davoodi, and Alonso-Terme (1998, 28) find that economic growth is inversely related to Transparency International's index of perceived corruption as well as the indicators used by Mauro. And Friedman et al. (2000) trace informality, tax evasion, and their attendant political and economic ills to a number of different indicators of bureaucratic inefficiency and corruption. "So widespread is the confidence in these findings," writes a justifiably cautious Seligson, "that international lending agencies have embarked upon major efforts to reduce corruption, conditioning many of their loans on formal, widespread efforts to clean it up" (2002, 410).

Nor are the international lending agencies alone. The Bush Administration partially limits access to the foreign aid provided under the auspices of the Millenium Challenge Account to countries that display superior governance—as measured by, *inter alia*, the "aggregate governance indicators" developed by Kaufmann et al. for the World Bank (*Economist* 2005; Radelet 2002, 2003). While Kaufmann holds that that governance fosters growth and not vice versa and asserts that "a country that improves its governance from a relatively low level to an average level could almost triple the per capita income of its population in the long term" (Kaufmann 2005, 41), he and his collaborators laudably admit that their indicators may be too blunt for policymaking purposes.[8] "In a ranking of 61 poor countries for which data were available in 2000–01," notes the *Economist*, "they could be 90% certain that Sudan and Burundi were correctly classified in the bottom half of the table. They could not be so sure of any of the other 28 countries that would fail to make the cut" (*Economist* 2005, 75).

We not only acknowledge and underscore the imprecision in Kaufmann's estimates but demonstrate that—within the limits imposed by his admittedly problematic but nonetheless increasingly popular measures—good governance is in all likelihood a consequence, rather than a cause, of economic growth and

[6]As obvious as these questions are, they have not attracted widespread scholarly attention. Kaufmann and Kraay (2002) and Kaufmann (2003–2004) are among the few analysts to pose these questions directly, finding that governance is a direct cause of development. Glaeser et al. (2004), however, suggest that good institutions are not nearly as important to growth as commonly thought, while Ritzen, Easterly, and Woolcock (2000) have suggested that the degree to which public institutions can be improved is highly constrained by societal factors.

[7]One of the best recent attempts is that of Acemoglu, Johnson, and Robinson (2001), who use the mortality rates of colonial settlers as an instrument for the quality of early political institutions. The intuition is that where mortality was low, higher levels of colonial immigration were possible, which promoted the development of bigger, more effective states. These early institutional advantages are then assumed to persist into the present era. Another approach, Kaufmann and Kraay (2002), relies on external information about measurement error in indicators of good governance to identify a system of equations linking governance to growth, and the reverse. This approach, which comes to quite different conclusions as Acemoglu, Johnson, and Robinson, relies on heroic assumptions about the nature of the errors in measurement, the degree to which contemporary measurements of governance are proxies for historical data on the quality of governance, and the unknown correlation between the errors in the system of equations.

[8]Kaufmann's estimated payoff to good governance has diminished of late. In 2003 he posited a 400% improvement in per capita income attendant upon a broadly similar improvement in governance (see Kaufmann cited in Francis 2003). Earlier he asserted that "halving the level of corruption" in Russia "would see per capita income, at least double, perhaps quadruple" (Kaufmann cited in Sweeney 1999).

the current effort to build "administrative competence" as part of a policy imperative is therefore at best insufficient and at worst misguided. Moreover, we suggest that there may be underlying political and social structures that can independently promote both effective state building and economic development, and until they are empirically investigated, and their effects estimated, we must remain cautious at best about any assertions of a *causal* linkage between governance and growth, however intuitively appealing it might be.

The Intuition. Our perspective builds on observations in the extensive qualitative literature linking public action and economic development. Many scholars have made the case that unusually high-quality public sector performance characterizes the polities of the newly industrialized countries (NICs) of East Asia. Indeed, scholars of all stripes, from developmentalists like Wade (1990) and Amsden (2001) to neoclassical economists at the World Bank (1993, 6) acknowledge that "government interventions resulted in higher and more equal growth than otherwise would have occurred" in the East Asian region. But these governments were not always particularly capable. The Kuomintang ruled mainland China through a combination of cronyism, clientelism, and naked force until 1949. It is hard to imagine that these same political leaders created a "developmental state" in Taiwan out of whole cloth a few short years thereafter. Similarly, the South Korean government of Syngman Rhee was known for its corrupt practices, economic malgovernance, and slow growth. The mere advent of a military coup in 1961 seems inadequate to explain the oft-asserted professionalization and modernization of the Korean state apparatus—and the wholesale modification of the developmental strategy and the achievement of world-beating economic growth rates over most of the subsequent 35 years. And indeed, recent evidence suggests that substantial problems of public probity and crony capitalism persisted throughout the long period of rapid economic development (Kang 2002). This forces one to ask whether development helped produce the developmental state almost as much as the developmental state impelled rapid economic development.

Similarly, quite a few of the countries currently among the most developed in the world were, during the period of their industrial takeoffs, clearly malgoverned and riddled with corruption. Glaeser and Shleifer (2001), for example, go to great pains to demonstrate the degree to which U.S. economic governance between the Civil War and the Roosevelt and Wilson administrations was shot through with cronyism and corruption, rendering corporate behavior almost immune to effective oversight. Indeed, the rise of regulatory agencies at the state and federal levels during the Progressive era was largely due to the overwhelming corruption of the judicial system, then the principal entity that governed economic practices. Nonetheless, during this period (from the 1860s to the 1900/1910s) the U.S. industrial economy underwent a dramatic and sustained expansion. And indeed, *in the wake of this development*, substantial improvements in the quality of governance were completed, including direct and responsible federal oversight of the money supply, banking, and interstate commerce; the professionalization of the civil service; and the regulation of trusts and monopolies. Similarly, rapid economic modernization in postwar Italy was possible almost in spite of, rather than on account of, an often corrupt, and typically unstable political system. Indeed, even as Italy remains a wealthy European nation, the headlines of its dailies continue to be dominated by charges of corruption—stunning for both their size and the upper reaches of government that they so frequently touch.

Our point is simple: Clean, effective government is desirable, but what is not so clear is whether it is an essential or even important antecedent of rapid economic growth—let alone whether it can be created through the administrative and judicial reforms most commonly recommended by donor governments and international financial institutions. Such reforms may in the end be essential, but they may also be ineffective in the absence of economic development or simply find their emergence blocked until underlying socioeconomic structures or sociopolitical interests are transformed. We also worry that popular measures of malgovernance are only partially adequate. Before we can with certainty estimate the strength (and causal direction) of the growth-governance linkage, we need measures of the latter uncontaminated by knowledge of antecedent economic performance or assumptions about economic policy choices.

We contend, in fact, that the record of political reform is far better in the places in which economic development has taken place—that is, political reform is more a consequence of economic reform than its cause. This does not imply that political development is an automatic consequence of economic expansion, but rather that political reforms are both more likely and more likely to succeed where such development has already taken place. This subtle point has profound consequences. It suggests that political modernization cannot be had on the cheap "merely" through the implementation of administrative and judicial

reforms—though these are certainly valuable in and of themselves. Instead, it may require ongoing efforts to undertake the hard and costly work of economic development—efforts that may well be impeded by government inefficiency but without which governance will not be improved.

Measurement: Are We Sure We Know How Good a Government Is?

To know whether good governance induces growth requires us to be able to measure the quality of public administration in a cross-nationally valid way. This is difficult enough, but it is made all the more so because operationalization begs the prior question, characterized by ongoing differences of opinion, of what government should (and should not) be doing in the first place. As a conceptual matter most economists—with some notable exceptions—subscribe to some variation of the maxim "he who governs best, governs least" (see Becker 1995; Krueger 1974; Shleifer and Vishny 1993). It is an approach that is dominant in the cross-national research.

But measuring state capacity in the manner most common among economists—in terms of what the state refrains from doing (regulating, taxing, stealing)—is neither easily nor necessarily profitably accomplished (see Hopkin 2002). Measurement typically relies, in whole or in part, on survey instruments—applied, alternatively, to foreign investors, domestic firms, or citizens. Questions seek to glean assessments of the national legal system, the level of "red tape," the speed of the permitting process, or the extent of corruption (see, e.g., Business Environment Risk Intelligence 2006; Chong and Calderón 2000; Mauro 1995; Transparency International [TI] 2004; World Economic Forum 2004). For Mauro, for example, the results "are taken to represent [international] investor's assessments of conditions in the country in question" (1995, 684). These approaches and other information have been incorporated into the ambitious metasurvey-based aggregate governance indicators developed by Kaufmann and colleagues at the World Bank (Kaufmann and Wei 1999; KKM 2003, 2005). This project has many commendable features and clearly represents the state of the art. That said, important questions remain.

Reliance on these sorts of surveys, in whole or in part, requires the assumption that the interests of investors (foreign and domestic) and the interests of the nation are essentially coterminous. But this is an exceedingly selective notion of state capacity, and efforts at measurement that hinge on surveys of businesspersons are thus likely to contain substantial biases. Why? To the extent that public bureaucracies *are* effective in imposing taxes or regulatory demands (e.g., securities and prudential banking regulations, labor laws, industrial performance standards, environmental controls, or antitrust actions), they are likely to be judged "burdensome" and "growth-inhibiting" by many businesspersons. By contrast, where such controls don't exist or are easily evaded, states will be judged less harshly by business elites. This introduces policy preferences into measures of governmental quality or effectiveness and thereby injects *systematic* bias into the measures to the extent that public policy mirrors or diverges from the interests of surveyed business elites. This is unfortunate, since good governance is in principle conceptually independent of policy choices—it means that public officials are willing and able to effectively implement policy choices, whatever they might be. The key here is thus *not* measurement error in the sense of signal-to-noise problems. Rather, it is systematic bias based on the policy preferences of vested interests that would make even perfectly reliable measurement of perceived levels of governance diverge from the actual underlying level of administrative competence.

But the problems do not end here. Surveys of businesspeople are riddled with potential sample selection problems. They systematically censor the opinions of former investors who did not succeed in the marketplace, or potential investors who were deterred from entering local markets by pervasive malgovernance or corruption itself, and thereby sample a very unrepresentative group of firms.[9] This is not easily remedied—it is generally impossible to identify, and impractical to interview, "potential" investors deterred by malgovernance and/or malfeasance from entering local markets. By contrast, investors who *are* competing successfully in the marketplace, and therefore show up in the surveys, may be doing so precisely because they are the beneficiaries of corruption and cronyism—and are therefore unlikely to report it accurately. And where malgovernance *is* effectively reported, this may well be because it is *not* pervasive enough to create sufficiently strong distortions in firm-level survival or investor behavior to induce selection bias. And thus in such contexts those who do not win from malfeasance can survive to report it! But how can we determine which situation obtains in a particular case?

[9]Hopkin (2002) notes that studies of corruption have also tended to select on the dependent variable, often not examining comparable cases in which corruption was less severe.

An additional problem that may bedevil not simply business surveys but all opinion data is the possibility that respondents' estimates of bureaucratic competence are colored by cultural blinders—i.e., people in different countries have different definitions and opinions of "corruption"—and recent economic performance (see Seligson 2006). A government that presides over a period of strong growth may be perceived by many respondents, ceteris paribus, as comparatively efficient and effective regardless of actual bureaucratic practice—especially in light of the aforementioned conventional wisdom regarding the nature of the growth-governance linkage. By way of contrast, a government that presides over crisis, like the ones that occurred in Korea and Argentina in the late 1990s and early 2000s, will almost certainly be perceived as more incompetent and corrupt—whether the depth or extent of malgovernance has actually changed (see, e.g., Seligson 2006, 385 on Argentina). This is particularly true for citizen surveys that perforce include principally respondents who have little direct basis on which to form judgments of the quality of public administration other than easily visible knowledge of economic or other basic performance measures. While growth rates and bureaucratic quality may be correlated in the very long term, since most scholars think institutions change only slowly and/or episodically (Evans and Rauch 1999), a valid survey-based measure of governance should not move in tight relationship to short-term changes in economic growth.[10]

Clearly, the most comprehensive source for cross-national measures of governance is the series of indicators developed by KKM (2003, 2005) at the World Bank.[11] Of the six principal governance indicators produced by KKM, only the measure of "government effectiveness" clearly attempts to capture the ability of the state to formulate and implement its goals. This they define, quite properly, as "the competence of the bureaucracy and the quality of public service delivery" (2005, 4). Two of the other indicators are measures of regime characteristics ("voice and accountability" and "political stability") that are not conterminous with governance, while the measure of "regulatory quality" is premised on the notion that minimal regulation and minimal barriers to trade and investment flows are optimal and is thus conflated with (controversial) policy prescriptions. Measures of the "rule of law" have useful data on the enforceability of private and government contracts and the costs and independence of the judicial system, but are similarly conflated with policy preferences over the structure of private property rights, and business-elite oriented questions about whether judicial action "interferes" with business.[12] Similarly, the measure of "corruption control" unfortunately combines survey results as to the presence of nepotism, cronyism, and bribe taking in government with questions about the "intrusiveness of the bureaucracy" or the "amount of red tape." But just as in the rule of law case, intrusiveness and red tape can be a sign of *either* effective or ineffective governance, depending on the content of the policies being enforced.

Finally, when it comes to evaluating the growth-governance linkage, the policy biases embedded in these measures become even more problematic. For example, one prominent school of thought has highlighted the importance of developmentalist policies and competent but interventionist bureaucracies for rapid economic development (e.g., Amsden 2001; Wade 1990). Those working in this context have pointed out, according to Amsden (2001), that such states are necessarily "disciplinary" of capitalists—something that survey measures of businesspeople's opinions are likely biased against. A simple example will illustrate the problem. In his classic study of the developmental state in Taiwan, Wade (1990) notes that Kuomintang officials compelled export-oriented North American electronics firms to source their inputs locally by, first, delaying their applications for import permits and, second, introducing them to capable local suppliers. In the qualitative case study

[10]It might be thought that the aforementioned work of Acemoglu, Johnson, and Robinson (2001) overcomes these objections by using data on settler mortality as an instrument for the quality of governance in contemporary polities. Nothing could be further from the truth. After all, the variable for which Acemoglu, Johnson, and Robinson instrument is a measure of "expropriation risk" as perceived by foreign investors (2001, 1377), a variable that suffers from all the same selection and perception problems identified above. Nor should it escape notice that Acemoglu, Johnson, and Robinson's instrument for expropriation risk—which, importantly, is *not* the same thing as state capacity—are the mortality rates of the biggest expropriators in history: the European colonists.

[11]Kaufmann, Kraay, and Mastruzzi construct a meta-indicator that aggregates a host of different measures, from firm, investor, and population surveys to expert and international organization assessments to come to their overall measurements of the quality of governance. The only other reasonably broad survey, that of Transparency International, is not as complete, incorporates fewer source inputs, and in 2001 chose to eliminate citizen survey data altogether (Lambsdorff 2001, 2).

[12]The problem is that government interference is often a symptom of *good* governance (e.g., when public action prevents negative externalities, inhibits monopolies, or draws investors into productive sectors under developmentalist policy regimes). At the same time it can foster or signal inefficiency, the prevalence of graft, or judicial capture by private agents.

literature, these actions are considered the essence of good government, for they generated additional value added and thereby deepened the country's industrial structure. But in constructing their own indicator of "government effectiveness," Kaufmann and his colleagues have explicitly equated the "quality of bureaucracy" with the absence of "red tape" and have quoted one of their source surveys to the extent that "the better the bureaucracy the quicker decisions are made and the more easily foreign investors can go about their business" (KKM 2003, 93). Taiwan, by this measure, was poorly governed. Of course the problem is that bureaucratic delay can indicate *either* malgovernance *or* an effective state that seeks to compel business to behave in ways consistent with the long-run national interest rather than short-run private profit. The insensitivity of the existing quantitative measures to this particular problem might explain why Taiwan and South Korea are ranked 32nd and 42nd, respectively, in terms of government effectiveness while being almost universally hailed in the qualitative literature for possessing unusually high-quality public administrations. The problem is potentially more severe in studies that use these measures to assess the relationship between free-market policies and the quality of governance—since the former will tend to foster the latter by design.

Nor do the problems stop there. The KKM measure also incorporates questions about the quality and reliability of public and quasi-public goods like infrastructure, schools, and telecommunications (KKM 2003, 93). We worry not that public and quasi-public goods are unimportant but that their quality and reliability are likely to (1) reflect policy decisions as well as institutional capacity and (2) have independent—and therefore statistically inseparable—effects on growth in any event. Is growth a product of the quality of public services or the volume of public investment? Unfortunately, questions like, "How problematic is transportation for the growth of your business?" (KKM 2003, 93) are unlikely to provide the answer.

Because of these serious potential biases as well as the incongruous results across the quantitative and qualitative evaluations of state capacity, it is very important that the validity of our quantitative indicators be carefully examined before they are used to support or refute hypotheses linking governance and growth. This is, of course, more easily said than done. Here we take three approaches to the validation of the governance measure: (1) Do repeated observations taken at different points in time correlate with each other? (2) Do alternative indicators of governmental performance correlate with each other? and (3) Can construct validity be established?

We should emphasize that while KKM have been quick to point out that their indicators of government effectiveness necessarily contain measurement error, this is not our principal worry. While random error in measurement is problematic, it is tractable. Indeed, with respect to this type of problem their aggregated measures are clearly state-of-the-art. Our concern is with potentially *systematic* errors that may result from selection problems, perceptual biases, and survey design and aggregation. While KKM have made much progress, we worry that the study of governance may to some extent still be characterized by what Klitgaard, Fedderke, and Akramov call "an explosion of measures, with little progress toward theoretical clarity or practical utility" (2005, 414).

Reliability. We begin by examining the stability of KKM's measure of government effectiveness across time. It has long been conceptually established that quality of governance is a feature of public administration that tends to change only very gradually over time. Indeed, Acemoglu, Johnson, and Robinson (2001) go so far as to suggest that differences in the quality of governance at the dawn of colonization between the sixteenth and nineteenth centuries are quite well associated with the character of contemporary political institutions. Evans and Rauch (1999) are comfortable with the far less heroic assumption that the quality of bureaucratic structures is effectively constant over periods of at least 20 years in length. By this standard, we propose a simple test: Do the measures of government effectiveness correlate with each other across the four observations available in the 1996–2004 period?

If the assumption that the underlying quality of public administration is constant over short periods of time is reasonable, then the Kaufmann data are effectively repeated observations of the same concept. That being the case, if the measure is reliable we would expect these repeated observations to be very highly correlated with each other. The results (available from the authors) show strong cross-temporal correlation—as would be expected of measures of a concept usually thought to be constant over short periods. The bivariate correlations vary in strength from .902 to .965. This gives us a sense that the KKM measures are picking up a consistent underlying concept. But is it governance?

Validity. This does not yet, however, address the validity of the concept—is "government effectiveness" really capturing (just) the quality of the public administration? To begin to assess whether in fact this is the

case, we examine whether this measure correlates with the next most widely employed indicator of bureaucratic quality, Transparency International's (TI) Corruption Perceptions Index. Data from 2000 are employed as earlier TI datasets are confined to a relatively smaller and disproportionately wealthy subset of countries, naturally overrepresenting cases at one end of the governance spectrum. That said, the TI data still cover fewer than half the number of countries available in the KKM dataset. Despite this, the measures are quite strongly correlated ($r = .922$). Similarly, the "country risk" measures from the International Country Risk Guide, another widely employed proxy for the quality of governance, is also quite strongly correlated with KKM's government effectiveness measure ($r = .821$ for the four periods). While these results are certainly comforting with respect to the validity of the KKM measure, they are far from definitive. All these measures are liable to suffer shared biases as a consequence of their underlying methodological similarities—a reliance on firm, investor, and/or citizen surveys and a conflation of indicators of policy choice and governance quality.

The results are far less felicitous, however, when compared with another measure of government effectiveness that is not constructed through the reliance on citizen or investor surveys. Evans and Rauch (1999; [ER]) have produced a strictly institutional measure of bureaucratic quality, which they call "Weberianness," for 35 middle- and lower-income countries in the mid-1990s. For us, the key difference is that the ER measures are not obviously subject to either contamination with indicators of policy choice or biases introduced by the perceptions or preferences of citizens or investors. The timing of their measures is also essentially the same as (especially the earlier) KKM measurements. The correlation between the ER and KKM measures ranges from .587 to .649. This is at best a modest relationship (given that they should be measures of the same concept), and it is consistent with our worry that the KKM measures, while capturing aspects of government effectiveness, are probably *also* capturing biases induced by the simultaneous incorporation of policy indicators and the misperceptions of the (potentially biased) survey respondents on whom they rely.

We proceed, however, using KKM's government effectiveness measure instead of either alternative. In addition to its greater popularity and growing policy relevance, it has two principal strengths that commend its use: it displays reasonable reliability and has much broader coverage, avoiding sample selection problems at the country-level. Such problems would loom large were the smaller and nonrandom TI or Evans and Rauch data sets used.

Our next task is to examine the convergent and discriminant validity of the KKM measure. Fortunately we have strong theoretical expectations we can use to structure this assessment. First, almost all analysts would expect government effectiveness and the level of development to be strongly correlated (though the direction of causality would be in dispute). Second, it is widely expected that levels of education prevailing in the adult population and the quality of the bureaucracy would be positively related (Rodrik 1994). Finally, we examine whether the size of the population is related to the quality of governance, controlling for wealth and education. This follows from the argument that, all else equal, larger societies are more complex and in principle more difficult to administer (Xin and Rudel 2004). These hypotheses, then, provide standards against which convergent validity can be assessed. By contrast, since most analysts consider governance quality to be substantially constant over relatively short periods of time (e.g., 20 years or less), we would have a strong prior for discriminant validity: government effectiveness should *not* vary with the rate of recent (antecedent) economic growth. Indeed, to the extent that it does, it is possible that perception-bias tied to economic performance is corrupting the measure of governance, or growth itself is improving governance even in the very short term.

Table 1 presents the results of a series of tests of both convergent and discriminant validity. In Models I through IV each biannual observation in the KKM data set is examined separately. Because the data are normalized to mean zero, standard deviation one on an annual basis, the year-to-year changes in governance score are not, in the strictest sense, directly interpretable, though they are clearly appropriate for cross-sectional analysis. Nevertheless, we include a pooled model for comparative purposes—as KKM (2005, 2) point out that there is no discernable year-to-year trend in the governance averages.[13] In all models

we find, as expected, a strong positive relationship between wealth and governance. Regardless of the model estimated, GDP/capita maintains a substantively and statistically important relationship to government effectiveness. But this is not the case with the

[13]Strictly speaking this is still not appropriate, since the data were not only mean centered but also set to a standard deviation of one for each year. To be perfectly valid one would have to assume that the original data in question were distributed similarly across each of the years.

TABLE 1 Convergent and Discriminant Validity: How Well Does "Government Effectiveness" Measure the Effectiveness of Government?

Dependent Variable:	I Government Effectiveness, 1996	II Government Effectiveness, 1998	III Government Effectiveness, 2000	IV Government Effectiveness, 2002	V Random Effects Pooled Analysis, 1996–2002
GDP per capita	.111***	.097***	.081***	.081***	.073***
	(.008)	(.009)	(.008)	(.008)	(.009)
Education	.010	.018	.018	.045	.075***
	(.023)	(.024)	(.024)	(.025)	(.027)
GDP growth rate$_{(t-1,\ t-2)}$	.063***	.072***	.068***	.046**	.031***
	(.014)	(.019)	(.013)	(.022)	(.007)
log (population)	−.031	−.044*	−.016	−.033	−.033
	(.024)	(.026)	(.025)	(.027)	(.022)
Constant	−.359	−.170	−.498	−.437	−.490
	(.392)	(.422)	(.431)	(.465)	(.383)
Year 1996					.181***
					(.039)
Year 1998					.151***
					(.037)
Year 2000					.082***
					(.026)
N	104	105	105	104	418
R^2	.87	.80	.81	.84	.81

Notes: GDP/capita expressed as thousands of U.S. dollars at purchasing power parity for the year in question. Education is measured as the average number of years of schooling in the over-15 population in 1990 (last year available). GDP growth rate is the two-year average rate of GDP per capita growth, for the two years prior to the measurement of government effectiveness.
Sources: GDP/capita, population, and growth rates from World Bank (2005); Education from Barro and Lee (1996). Government Effectiveness from KKM (2005). For details, see the appendix (available online at http://www.journalofpolitics.org).
***$p < .01$; **$p < .05$; *$p < .10$.

educational attainment in the population. Here, while all the parameter estimates are appropriately signed, none achieves statistical significance, save for the pooled model (V).[14] While this limited relationship is a cause for concern, it is certainly not a definitive test of the validity of the government effectiveness measure. Population educational attainment is measured approximately a decade before the KKM governance data, the latest time period available. All else equal, a larger population also seems related to a lower governance score, though again these parameter estimates do not achieve even minimal statistical significance except in one case (Model II).

The test of discriminant validity is more troubling. If perception bias is a real problem in survey-based measures of bureaucratic quality, then we should see a strong relationship between *antecedent* economic performance and the governance quality measure. If on the other hand the KKM measure does effectively capture the fairly stable underlying quality of the public administration, this should be largely unaffected by short-term fluctuations in growth—the quality of governance should, after all, not simply follow the business cycle. Here the results are quite troubling. Across all of the models (I–V), antecedent economic growth (the average of the two years prior to the governance measure) is a strong predictor of government effectiveness. It seems that either economic performance induces biases in perceived governance quality, or we must believe that growth almost instantaneously induces improvements in governance. Whether this is really the case depends in part on whether one believes that economic improvements can be translated into institutional improvements in the very short run. As a whole these results raise the unfortunate possibility that while the KKM governance measure partially captures the underlying concept, at the same time it may also be substantially contaminated by respondents' perceptions of immediate economic conditions or biases that are products of sample selection.

Does Good Government Cause Growth?

Recent scholarship has emphasized the importance of good governance for economic performance. Mauro has gone so far as to declare that "a consensus seems to have emerged that corruption and other aspects of poor governance and weak institutions have substantial, adverse effects on economic growth" (2004, 1). More typical are efforts like those of Kaufmann (2003–2004) and Kaufmann and Kraay (2002) to explore the causal linkage between good governance and growth. These approaches have, however, been almost entirely cross-sectional in nature, utilizing either simple OLS or instrumental variables approaches. Kaufmann and Kraay (2002, 22) depart from this somewhat, by introducing a novel simultaneous equations model to assess the direction of the causal relationship, relying on a series of assumptions about nonsample information to achieve identification. Neither the instrumental variables nor simultaneous equations approaches are entirely satisfying as there is much disagreement as to whether appropriate instruments exist (see Durlauf, Johnson, and Temple 2005; Frankel et al. 2003; Glaeser et al. 2004), and the assumptions about measurement error necessary for the identification of the simultaneous equations model are implausible in the face of the biases discussed here.

We suggest a simple alternative. While Kaufmann's measure has been shown to have statistical power in cross-sectional research (Kaufmann, Kraay, and Zoido-Lobatón 1999), the true test of his theory must be longitudinal (Lieberson 1985). Only then can we be confident that the survey responses used in the construction of the measures were uninfluenced, for example, by the recent growth history of the country in question. And if a measure of state capacity is to be useful, it must be because it can help tell us whether we can expect, ceteris paribus, *future* growth in that country. But can the Kaufmann measures predict *future* growth?

The Kaufmann data set is of very recent vintage, and it therefore circumscribes our ability to carry out all but the most rudimentary of analyses. Five separate iterations of this indicator have been produced, biannually between 1996 and 2004. Since we are interested in predictive power, the last two sets of measures are not helpful, for enough data on cross-national growth rates are not available after 2003. We saw above that antecedent growth was quite predictive of scores on the quality of government measure. But if it is useful for policy purposes, the KKM measure must itself predict future growth (the two years after the taking of the KKM measure). In Table 2 we examine the relationship between government effectiveness and future growth in a very simplified model. Each model also controls for the level of economic development, as it is usually hypothesized that wealthier countries are not capable of as rapid rates of growth as the less developed—and thus omission of this variable might

[14]It might be that this is because GDP/capita and years of schooling in the adult (over age 15) population are strongly related. The correlation between these two variables is r = .75.

Table 2 Does Government Effectiveness Predict Subsequent Growth?

Dependent Variable: GDP/capita growth	I 1996	II 1998	III 2000
GDP/capita	.082	.086	−.023
	(.149)	(.074)	(.048)
Government	−.337	.005	−.118
Effectiveness	(1.436)	(.777)	(.608)
Constant	1.473*	1.474***	1.819***
	(.856)	(.502)	(.418)
N	163	164	163
R^2	.006	.04	.007

Source: GDP/capita and growth rates from World Bank (2005). Government Effectiveness from KKM (2005).
***$p < .01$; **$p < .05$; *$p < .10$.

lead to a spurious (negative) association between government effectiveness and growth since the former is so tightly correlated with wealth. None of the panels provides support for the hypothesis that governance is a useful predictor of *future* economic growth, at least with the limited two-year time horizon that we employ. Indeed, no relationship at all appears in the data. In the online appendix, to check for robustness, we also estimate a series of alternative basic models, which in no case produce a positive or significant association between government effectiveness and subsequent growth.

Still, this is only a very preliminary examination. It is well known that economic growth responds to a series of other factors, whose omission could be affecting the results we present. Investment levels and the human capital stock are, after all, quite likely to be correlated with the quality of public administration. Our measure of the former is the level of investment (gross fixed capital formation) relative to GDP in the antecedent year, while human capital is measured as the average years of schooling in the adult (over age 15) population in 1990.[15] In addition, controls for the logarithm of the inflation rate in the antecedent year are included to capture the effects of short-term crises on growth rates.[16] Regional dummy variables are also included in models I to III to try to capture the effects of unmeasured regional heterogeneity. In model IV we present a pooled analysis, which vastly increases our analytic leverage. This permits us to use country and year dummy variables to account for underlying national characteristics and time-bound international shocks that affect growth performance.

Table 3 presents the results of this analysis, which at first blush suggest that government effectiveness bears at best an uncertain relationship to subsequent levels of economic growth. Only for model II is the coefficient substantively fairly large and statistically significant at conventional levels. In model I, government effectiveness retains a positive relationship to subsequent growth, but its estimated effect is not statistically significant. In model III, the estimate falls far short of conventional statistical significance. The pooled model is the most troubling, however, for proponents of the governance-growth linkage. Here, the only model in which country-specific effects can be controlled—which is crucial as countries are well-known to have distinct underlying long-term "normal" growth rates that reflect their specific individual institutional and political conditions—in fact produces a *negative* (but insignificant) coefficient on the estimate of the relationship between governance and growth.

Since the government effectiveness measure is normalized to a mean of zero and a standard deviation of one, the effect estimate can be understood as the increase in the two-year average growth rate for a standard deviation increase in this indicator. The level of wealth has the conventional negative relationship with growth rates—it is widely assumed that poorer economies are able to grow at higher rates than wealthier ones. Neither investment levels nor human capital (education) have a consistent relationship to short-term growth in most of these models, though this may simply reflect collinearity problems as they are usually correlated with each other and the level of economic development.[17] Finally, crisis, at least as signaled by inflation, also does not have a statistically significant relationship to growth. This, however, may simply be an artifact of the tendency for reductions in growth to come as a consequence of stabilization efforts, not inflation per se.

[15]This is the last year for which data are available. Alternative measures are more problematic. Literacy rates suffer ceiling effects and enrollment rates measure *potential* human capital.

[16]The inflation rate is maldistributed on the right-hand side—using it would render a few hyper-inflationary cases far too important in the estimation. Logarithms of inflation rates that are occasionally negative, or positive but close to zero, are also problematic as the former are undefined and the latter will produce very large negative numbers. We therefore recode inflation rates less than 1% as 1%, so that their logarithm would be zero.

[17]As a robustness check, the pooled model was reestimated three times, serially removing either controls for gdp/capita, investment, or education in order to make sure that the coefficient on government effectiveness was unaffected by collinearity. In no case did doing so render the government effectiveness coefficient statistically significant, nor did its sign change to match conventional expectations.

TABLE 3 Government Effectiveness in a Basic Growth Model (Dependent Variable: Average Rate of Growth of GDP/Capita Over the Two Years Subsequent to the Measurement of Government Effectiveness)

	I 1996	II 1998	III 2000	IV Pooled Model
GDP/capita	−.166*	−.151**	−.211***	−.653***
	(.098)	(.065)	(.062)	(.234)
Government Effectiveness	1.372	1.133**	.583	−.516
	(.829)	(.534)	(.519)	(1.429)
Investment	−.053	.132***	.153***	−.173**
	(.051)	(.048)	(.053)	(.085)
Education	.017	.080	.216	N/A
	(.220)	(.176)	(.159)	
Log (inflation)	−.567	−.295	−.176	.101
	(.403)	(.281)	(.305)	(.561)
Africa	−1.687	−2.386**	−2.185	
	(1.549)	(1.151)	(1.289)	
Latin America	−.076	−3.843***	−4.595***	
	(1.359)	(1.075)	(1.309)	
Asia & Oceania	−2.836***	−.859	−2.181**	
	(1.074)	(1.081)	(1.028)	
Europe	−.321	−.895	−.407	
	(.823)	(.568)	(.580)	
Middle East	−1.433	−3.378***	−2.810**	
	(1.375)	(1.263)	(1.220)	
Year 1996				−.387
				(.799)
Year 1998				.466
				(.519)
Country Fixed Effects				[suppressed]
Constant	5.970***	2.355	1.203	−13.725*
	(2.079)	(1.758)	(1.879)	(7.035)
N	105	103	103	311
R^2	.187	.402	.341	.601

Notes: Pooled model estimated with robust standard errors, assuming clustering by country. Estimated in Stata 9 using the xtreg command. Model IV is effectively a fixed effects regression, and when estimated thusly the coefficient estimates are identical, save Schooling which is omitted as constant within all units. Similarly, an estimation of Model IV as in the table omitting the schooling variable results in an even more negative (but still insignificant) estimate of the effect of government effectiveness on future growth. The coefficient on education is suppressed as it is not time-varying, and thus not meaningful in what is effectively a fixed-effects specification.
Sources: GDP/capita at ppp, inflation, and investment (gross fixed capital formation as a share of GDP) and growth rates from World Bank (2005). Education from Barro and Lee (1996). Government Effectiveness from KKM (2005). Inflation rates less than 1% recoded to equal 1% before taking the logarithm.
***$p < .01$; **$p < .05$; *$p < .10$.

This is not, however, sufficient to sustain the oft-asserted notion that growth and governance are linked in a reciprocal and self-reinforcing relationship. The problem is that, as we saw in Table 1, governance is very tightly correlated to antecedent economic growth rates, which raises real questions as to whether perception biases are contaminating the measure. It is also well known that growth rates are serially correlated—the unmeasured factors making growth rates especially high (or low) in a particular year are likely to persist into subsequent periods. As a consequence, a more valid test of the linkage between governance and growth would try to control out that portion of the governance measure that is really due to a correlation with preceding rates of growth, and leave us with a much purer measure of institutional capacity.

In Table 4 we present the results of an analysis that attempts to do precisely this. Here we replicate the analysis of Table 4, but include an additional control

TABLE 4 Government Effectiveness and Growth, Controlling for Inertial Effects (Dependent Variable: Average Rate of Growth of GDP/Capita Over the Two Years Subsequent to Measurement of Government Effectiveness)

	I 1996	II 1998	III 2000	IV Pooled Model
GDP/capita	−.044	−.135*	−.152**	−.687***
	(.123)	(.068)	(.061)	(.241)
Government Effectiveness	.535	.932	−.179	−.784
	(1.066)	(.584)	(.521)	(1.504)
Investment	−.081*	.092	.130***	−.182*
	(.042)	(.058)	(.041)	(.094)
Education	−.021	.105	.246*	N/A
	(.227)	(.178)	(.137)	
Lagged GDP/capita growth (t-1, t-2)	.274*	.168	.323***	.058
	(.161)	(.163)	(.087)	(.132)
Log (inflation)	−.472	−.282	−.098	.088
	(.391)	(.301)	(.316)	(.564)
Africa	−.431	−2.337*	−1.180	
	(1.465)	(1.120)	(1.295)	
Latin America	.707	−3.700***	−3.549***	
	(1.312)	(1.178)	(1.278)	
Asia & Oceania	−2.071**	−.742	−.943	
	(1.030)	(1.130)	(.984)	
Europe	−.014	−.744	.012	
	(.737)	(.618)	(.544)	
Middle East	−.221	−2.942**	−1.653	
	(1.408)	(1.305)	(1.258)	
Year 1996				−.499
				(.807)
Year 1998				.391
				(.506)
Country Fixed Effects				[suppressed]
Constant	4.180*	2.426	.333	−14.205**
	(2.388)	(1.768)	(1.857)	(6.954)
N	104	103	103	310
R^2	.234	.419	.435	.603

Sources and Notes: see Table 3. Model IV was subjected to the same robustness checks as in Table 3.
***$p < .01$; **$p < .05$; *$p < .10$.

for antecedent rates of economic growth.[18] In no case is the coefficient on this variable close enough to unity to signal problems. Indeed, from the perspective of our argument, this is a conservative specification, as we are not simply incorporating the immediate lag of the dependent variable, but rather an average of the two periods that antedate it by a year. The results reinforce the doubts generated by the analysis in Table 3. Once inertial effects of growth are controlled for, government effectiveness is in no instance related to subsequent rates of economic growth. Moreover, in two of the models—III and IV—the sign of the estimate is indeed negative (though insignificant). The controls for wealth, human capital, and investment behave much as in Table 3.

Where does this leave us? We are still far from a definitive statement as to the relationship between good governance and growth. That said, several principal findings are apparent. First, it is likely that the KKM governance measure, while capturing important

[18]Antecedent growth is the average of the two years prior to the time of the government effectiveness measure. The dependent variable refers to the average of the two years *after* the taking of the government effectiveness measure. A gap of a year helps reduce the potential for bias that can be introduced by including a lagged dependent variable—the temporal separation helps to reduce the likelihood that this included variable is correlated with the error term.

aspects of the institutional quality of the public bureaucracy, is also contaminated by perception and/or selection biases. The strong linkage between reported government effectiveness and antecedent rates of economic growth suggests that respondents may, at least in part, be basing their assessments on this performance criterion rather than on the much more stable underlying features of the institutional organization of the state. This is critical to the measure insofar as these surveys form a component of KKM's index.

Second, insofar as the KKM measure is valid, we find only tepid support for the notion that improvements in governance lead directly to improvements in the short-run rate of growth. This is not to say that malgovernance is a good thing—nowhere do we find meaningful evidence that lower government effectiveness predicts higher rates of growth. But it does undermine the notion that improvements in public administration alone will improve *subsequent* economic performance. It is quite possible that other analyses that have found a strong such link do so because they are cross-sectional in design. In that context, the perception bias partly embedded in the measure of government effectiveness is likely to create a spurious correlation with growth rates. Our longitudinal analysis, while hardly sufficient to establish the appropriate causal direction, does cast doubt on the "virtuous cycle" assumption that is prevalent in the literature.[19] It also raises the possibility that economic performance can be improved even in malgoverned polities if "growth oriented" economic policies are implemented—even if they are "leaky" in terms of resource diversion. This may be because the economic losses entailed by malgovernance are not catastrophic relative to the gains to be had from policy improvement or because growth itself subsequently leads to the improvement of the public administration, providing in essence an intertemporal positive externality. If anything, it raises red flags about the current effort to condition international aid on the quality of governance.

Where Do We Go From Here?

This paper departed from two simple questions: Does good governance cause growth? Does growth improve governance? We also raised but did not explore the possibility that the widely heralded cross-sectional correlation between growth and governance is a largely spurious result brought about by underlying factors that promote, independently, both state building and economic development.

Lest the reader think we're attacking a straw man, note the growing academic and popular sense that "bad government is the single most important cause of failure" in the developing world (Wolf 2005; see also Castañeda 2003). Some observers go so far as to portray good government as a "necessary" precondition of economic development (see, e.g., M'Dhaffar quoted in Kim et al. 2005). And almost all parties acknowledge and underscore the centrality of governance to development (Francis 2003). "There has been a sea change in the past seven or eight years in awareness of the issue," suggests Kaufmann (quoted in Francis 2003, 16), and his readily available indicators are at least partially responsible. "Economists can now prove the enormous cost of corruption," writes David Francis of the Christian Science Monitor, and disseminate their findings over the Internet. "The World Bank site on corruption gets some 500,000 visitors a month," he concludes, "half from developing countries" (Francis 2003, 16).

Our results suggest that the data and conclusions found on the World Bank site—at least with respect to government effectiveness—are at best partial and at worst misleading, however, for we are at the beginning—rather than the end—of our efforts to unpack the complicated relationship between growth and governance. As a next step, we believe, we need better measures of governance, particularly ones that feature a much wider historical sweep and do not rely on surveys that embed perceptual and policy biases. Since many consider the effects of governance to be perceptible only over the relatively long term, it behooves us to find direct measures of governance that can be found for long historical periods. This would allow us to avoid either the assumption that institutional quality (or the global hierarchy of the same) is relatively constant over centuries (Acemoglu, Johnson, and Robinson 2001) or to project backwards over decades the results of contemporary analyses (Evans and Rauch 1999).

The literature on "democratization" provides a model. The Polity IV data set maintained by Marshall and Jaggers at the University of Maryland relies upon neither an overly broad definition of democracy nor a biased sample of respondents but instead employs disinterested expert evaluations of the narrowly institutional features of political regimes. It extends

[19]This is an enormously complicated task. Unless suitable instruments can be found—and the task has so far proven difficult indeed—we must rely on alternative approaches that are at best suggestive. Our approach has been to use a longitudinal analysis to try to gain some leverage on the direction of the causal processes. It is necessarily only a first step.

backward to 1800, includes annual observations, and is updated continuously. If academics and policymakers really believe that the impact of good governance is as profound as their scholarship and policy choices suggest, and are willing to put their increasingly consequential beliefs to the test, an equally comprehensive data collection effort would appear to be more than worthy of national or international support.[20]

The second step is to take seriously the underlying social and political dynamics that could potentially explain away the assumed causal connection between growth and governance. This is fertile but comparatively unplowed terrain. But the disjuncture between long-standing approaches to the understanding of state building—that have emphasized structural features of the economy or the international system such as resource wealth or strategic conflict—and studies of governance that have assumed that the improvement of public administration is largely a function of easily changed legal structures begs questions we must answer. Similarly, in the qualitative literature on East Asian development, which almost always emphasizes state capacity and "market governance" as key predictors of world-beating growth rates, underlying structural factors are often mentioned but not systematically explored. It is usually noted that these societies have unusually high levels of educational attainment, unusual social equality, or have radically transformed agrarian social structure and property rights—oftentimes *prior* to building administrative capacity.[21] But all of these factors could quite plausibly be directly related both to economic development and the building of strong states. Equally suggestive are the variations in the quality of governance and level of development even within a single polity. Consider the United States—effective governance and higher levels of development map quite directly onto long-run structural features of our society and economy. Why, for example, is the former plantation South persistently underdeveloped and malgoverned relative to the North and Midwest where more egalitarian distributions of property and an absence of chattel slavery prevailed (Schrank 2004)?

The balance of the evidence available to date leaves us with two imperfect conclusions. Either we cannot reasonably conclude that improvements in governance produce meaningful increases in the rate of economic growth, or the absence of such an observed connection implies that our conceptualization and measurement of governance is as of yet quite imperfect. We remain agnostic as to which (or perhaps both?) is true, but have sought to make the case that the oft-asserted connection between growth and governance lies on exceedingly shaky empirical pilings. At the same time, potentially flawed indicators of governance quality are being utilized by policy makers to condition development aid and shape development efforts. But until we know more about what is (and is not) malgovernance, and the process by which it can be cured, such conditionality may do more harm than good.

[20]Evans and Rauch have made a noble effort by devising an institutional measure of administrative capacity in low to middle income countries in the mid-1990s. But funding limitations restricted their effort to 35 countries and a single time point and to therefore regress *prior* growth rates on *current* administrative capacity under the explicit assumption that the latter variable is sticky over time. It is telling in this regard that the bivariate correlation between their more narrowly institutional measure of "Weberianness" and the measure we have examined (KKM) is about .6 for the 35 available countries—and that the relationship between the Evans and Rauch measure and *subsequent* growth rates is nonetheless insignificant. While we believe that a more comprehensive version of the ER measurement strategy would be more fruitful than the KKM approach, we are by no means convinced that it would reveal a causal relationship leading from governance to growth. Again: the opposite may in fact be the case.

[21]Nor is it obvious that radical social reform requires East Asian-style bureaucracy to be effective. See, e.g., Turits (2003) for a fascinating account of successful land reform under the patrimonial Trujillo regime in the Dominican Republic.

Acknowledgments

We would like to thank Sarah Brooks, Jonathan Hopkin, Luke Keele, Cheol-Sung Lee, Kenneth Shadlen, the participants in The Ohio State University comparative politics research workshop, and panelists and audience members at the 2004 annual meeting of the American Political Science Association for their comments and advice.

Manuscript submitted 7 October 2005
Manuscript accepted for publication 6 May 2006

References

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review* 91 (5): 1369–401.

Amsden, Alice. 2001. *The Rise of "the Rest": Challenges to the West from Late-Industrializing Economies*. Oxford: Oxford University Press.

Barro, Robert J., and Jon Wha Lee. 1996. "International Measures of Schooling Years and Schooling Quality." *The American Economic Review* 82 (2): 218–23.

Becker, Gary. 1995. "If You Want to Cut Corruption, Cut Government." *Business Week*, December 11.

Brautigam, Deborah. 1992. "Governance, Economy, and Foreign Aid." *Studies in Comparative International Development* 27 (3): 3–26.

Business Environment Risk Intelligence. http://www.beri.com (May 1, 2006).

Caselli, Francesco, and Massimo Morelli. 2003. "Bad Politicians." *Journal of Public Economics* 88: 759–82.

Castañeda, Jorge. 2003. "The Forgotten Relationship." *Foreign Affairs* 82 (3): 67–81.

Chong, Alberto, and César Calderón. 2000. "Causality and Feedback between Institutional Measures and Economic Growth." *Economics and Politics* 12 (1): 69–81.

Durlauf, Steven, Paul Johnson, and Jonathan Temple. 2005. "Growth Econometrics." In *Handbook of Economic Growth*, eds. Philippe Aghion and Steven Durlauf. Amsterdam: Elsevier, pp. 555–677.

Economist. 2005. "A Choosier Approach to Aid." April 23.

Evans, Peter, and James Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth." *American Sociological Review* 64 (5): 748–65.

Francis, David. 2003. "Why Fighting Corruption Helps the Poor." *Christian Science Monitor*, 13 November.

Frankel, Jeffrey, Nancy Birdsall, Jeffrey Sachs, and Guillermo Ortiz. 2003. "Panel Discussion—Promoting Better National Institutions: The Role of the IMF." *IMF Staff Papers* 50: 21–40.

Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobatón. 2000. "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries." *Journal of Public Economics* 77 (3): 459–93.

Geddes, Barbara. 1994. *Politician's Dilemma: Building State Capacity in Latin America* Berkeley: University of California Press.

Gerring, John, and Strom Thacker. 2004. "Political Institutions and Corruption: The Role of Unitarism and Parliamentarism." *British Journal of Political Science* 34 (2): 295–330.

Glaeser, Edward, and Andrei Shleifer. 2001. "The Rise of the Regulatory State." *NBER Working Paper No. 8650*: 1–39.

Glaeser, Edward, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *National Bureau of Economic Research Working Paper*. No. 10568: 1–54.

Gupta, Sanjeev, Hamid Davoodi, and Rosa Alonso-Terme. 1998. "Does Corruption Affect Income Inequality and Poverty?" *IMF Working Papers 98/76*. Washington: IMF.

Hopkin, Jonathan. 2002. "States, Markets and Corruption: A Review of Some Recent Literature." *Review of International Political Economy* 9 (3): 574–90.

Huntington, Samuel. 1968. *Political Order in Changing Societies*. New Haven: Yale University Press.

Jomo, K. S. 2000. "Economic Considerations for a Renewed Nationalism." *Journal of Contemporary Asia* 30 (3): 338–69.

Kang, David. 2002. *Crony Capitalism: Corruption and Development in South Korea and the Philippines*. Cambridge: Cambridge University Press.

Kaufmann, Daniel. 2003–2004. "Governance Redux: The Empirical Challenge" *Global Competitiveness Report*. Geneva: World Economic Forum.

Kaufmann, Daniel. 2005. "10 Myths about Governance and Corruption." *Finance and Development* 42 (3): 41–3.

Kaufmann, Daniel, and Aart Kraay. 2002. "Growth without Governance." *Economía* 3 (1): 169–215.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2003. "Governance Matters III: Governance Indicators for 1996–2002." Typescript. World Bank.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005. "Governance Matters IV: Governance Indicators for 1996–2004." *World Bank Policy Research Paper*. No. 3630: 1–60. Associated online data set available at: http://www.worldbank.org/wbi/governance/govdata/ (May 1, 2006).

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999. "Governance Matters." *World Bank Policy Research Working Paper*. No. 2196: 1–61.

Kaufmann, Daniel, and Shang-Jin Wei. 1999. "Does 'Grease Money' Speed up the Wheels of Commerce?" *NBER Working Papers*. No. 7093: 1–29.

Kim, Pan Suk, John Halligan, Namshin Cho, Cheol H. Oh, and Angela Eikenberry. 2005. "Toward Participatory and Transparent Governance: Report on the Sixth Global Forum on Reinventing Government." *Public Administration Review* 65 (6): 646–54.

Klitgaard, Robert, Johannes Fedderke, and Kamil Akramov. 2005. "Choosing and Using Performance Criteria." In *High Performance Government: Structure, Leadership, Incentives*, eds. Robert Klitgaard and Paul Light. Santa Monica, CA: RAND Corporation.

Krueger, Anne. 1974. "The Political Economy of the Rent-Seeking Society." *The American Economic Review* 64 (3): 291–303.

Lambsdorff, Johann. 2001. *Framework Document: Background Paper to the 2001 Corruption Perceptions Index*. Göttingen, Germany: Göttingen University and Transparency International.

Leff, Nathaniel. 1964. "Economic Development through Bureaucratic Corruption." *American Behavioral Scientist* 8 (2): 8–14.

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Lui, Francis. 1985. "An Equilibrium Model of Queuing and Bribery." *Journal of Political Economy* 95 (4): 760–81.

Mauro, Paolo. 2004. "The Persistence of Corruption and Slow Economic Growth." *IMF Staff Papers* 51: 1.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110 (3): 681–712.

Radelet, Steven. 2002. "Qualifying for the Millennium Challenge Account." Typescript. Center for Global Development.

Radelet, Steven. 2003. "Bush and Foreign Aid." *Foreign Affairs* 82 (5): 104–17.

Rashid, Salim. 1981. "Public Utilities in Egalitarian LDC's: The Role of Bribery in Achieving Pareto Efficiency." *Kyklos* 34, Fasc. 3: 448–60.

Reynolds, Lloyd. 1983. "The Spread of Economic Growth to the Third World: 1850–1980." *Journal of Economic Literature* 21 (3): 941–80.

Riedel, James. 1988. "Economic Development in East Asia: Doing What Comes Naturally?" In *Achieving Industrialisation in Asia*, ed. Helen Hughes. Cambridge: Cambridge University Press, pp. 1–38.

Ritzen, Jo, William Easterly, and Michael Woolcock. 2000. "On 'Good' Politicians and 'Bad' Policies: Social Cohesion, Institutions, and Growth." *World Bank Policy Research Working Paper No. 2448*.

Rodrik, Dani. 1994. "King Kong Meets Godzilla: The World Bank and the East Asian Miracle." In *Miracle or Design? Lessons from the East Asian Experience*, ed. Albert Fishlow. Washington, DC: Overseas Development Council.

Rodrik, Dani. 2005. "Why We Learn Nothing from Regressing Economic Growth on Policies." Typescript.

Rose-Ackerman, Susan. 1999. *Corruption and Government: Causes, Consequences, and Reform*. Cambridge: Cambridge University Press.

Schrank, Andrew. 2004. "Reconsidering the Resource Curse: Selection Bias, Measurement Error, and Omitted Variables." Presented at the annual meeting of the American Political Science Association.

Seligson, Mitchell. 2002. "The Impact of Corruption on Regime Legitimacy: A Comparative Study of Four Latin American Countries." *Journal of Politics* 64 (2): 408–33.

Seligson, Mitchell. 2006. "The Measurement and Impact of Corruption Victimization: Survey Evidence from Latin America." *World Development* 34 (2): 381–404.

Shleifer, Andrei, and Robert Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108 (3): 599–617.

Stern, Nicholas. 1989. "The Economics of Development: A Survey." *Economic Journal* 99 (397): 597–685.

Sweeney, Paul. 1999. "The World Bank Battles the Cancer of Corruption." *Global Finance* 13 (10): 110–14.

Transparency International. 2004. http://www.transparency.org (May 1, 2006).

Turits, Richard. 2003. *Foundations of Despotism: Peasants, the Trujillo Regime, and Modernity in Dominican History*. Stanford, CA: Stanford University Press.

Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton: Princeton University Press.

Weingast, Barry. 1995. "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development." *Journal of Law, Economics, and Organization* 11 (1): 1–31.

Wolf, Martin. 2005. "Aid Is Well Worth Trying." *Financial Times*. 5 July.

World Bank. 1993. *The East Asian Miracle: Economic Growth and Public Policy*. Oxford: Oxford University Press for the World Bank.

World Bank. 2005. *World Development Indicators* CD/ROM. Washington D.C.: World Bank.

World Economic Forum. 2004. *Global Competitiveness Report*. Geneva: World Economic Forum. Also at http://www.weforum.org/en/initiatives/gcp/index.htm (December 1, 2006).

Xin, Xiaohui, and Thomas K. Rudel. 2004. "The Context for Political Corruption: A Cross-National Analysis." *Social Science Quarterly* 85 (2): 294–309.

Marcus J. Kurtz is associate professor of political science, The Ohio State University, Columbus, OH 43210. Andrew Schrank is assistant professor of sociology, University of New Mexico, Albuquerque, NM 87131-0001.

Growth and Governance: A Reply

Daniel Kaufmann The World Bank
Aart Kraay The World Bank
Massimo Mastruzzi The World Bank

In this issue of the *Journal of Politics*, Marcus Kurtz and Andrew Schrank (hereafter KS) offer a sweeping critique of the existing literature on governance and growth. They argue that perceptions-based cross-country measures of governance, and in particular those we have constructed in our ongoing work,[1] are fatally flawed. They also produce empirical evidence which they claim shows that perceptions of governance are driven by short-term growth performance. Finally they argue that there is little convincing evidence that good governance spurs growth.

In our response we show that these claims are unsubstantiated. We first show that their claims of biases in perceptions-based measures of governance are speculative and, to the extent that they are falsifiable, do not withstand empirical scrutiny. We next show that the empirical evidence in support of their claim that governance perceptions respond to short-run growth is both statistically fragile and conceptually flawed. Finally we dismiss their empirical work on the effects of governance on growth, which we argue is far removed from the best-practice frontier in cross-country growth empirics. We instead briefly describe the rich body of recent work in the economics literature that has documented a sizeable long-run effect of governance on growth.

Getting Concepts Straight

Before delving into the details of our response, we first note that KS's definition of governance is in our view convoluted and ultimately too narrow. In their opening sentence they refer broadly to "political corruption and malgovernance." By the next paragraph they have shifted to a much narrower concept of the "quality of public administration." Subsequently, the definition changes again to "the ability of the state to formulate and implement its goals." KS emerge from these shifting definitional sands to focus on just one of the six measures of governance that we construct, "Government Effectiveness."

While it is easy to get into endless terminological tussles over what governance is, here we simply want to make the point that KS focus narrowly on just one of our specific measures of governance and in our view inappropriately ignore other dimensions of governance that have received much more attention in the empirical literature on governance and growth. As we discuss further below, leading papers in this literature tend to focus on a more basic notion of governance going back to the seminal work of Douglas North: the norms of limited government that protect private property from predation by the state. This concept is much more closely related to our measures of Rule of Law and Control of Corruption, as well as several other indicators of these concepts.

We do not want to make too big a deal of this conceptual distinction because in the end these aspects of governance tend to be quite highly correlated across countries. There are not many countries where corruption is high yet the public sector manages to provide public services effectively, which is what our Government Effectiveness measure captures. Yet for the sake of conceptual clarity, and for the sake of placing this paper in the context of the existing

[1]These governance indicators capture six dimensions of institutional quality: Voice and Accountability, Political Instability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. The indicators cover over 200 countries and are available for 1996, 1998, 2000, and annually since 2002. The latest update of the governance indicators is described in Kaufmann, Kraay, and Mastruzzi (2006b), and the data and a web-based graphical interface are available at http://www.govindicators.org. Our earlier papers documenting these governance indicators are Kaufmann, Kraay, and Zoido-Lobatón (1999a, 1999b, 2002) and Kaufmann, Kraay, and Mastruzzi (2004, 2006a, 2006b).

The Journal of Politics, Vol. 69, No. 2, May 2007, pp. 555–562

 ISSN 0022-3816

empirical literature, it is important to note that KS's focus on this one specific measure of governance is very narrow indeed.

He Says, She Says: Are Perceptions Measures Really Biased?

KS begin their discussion with a series of assertions that perceptions-based measures of governance are biased in various ways. KS first argue that businesspeople's views of what good governance is might be very different from other views more broadly reflective of the public interest. In short, they argue, businesspeople like low taxes and less regulation, while the public good demands reasonable taxation and appropriate regulation. Estimates of governance based on the perceptions of businesspeople will therefore necessarily be biased.[2]

In response, we note first that our six aggregate governance indicators rely on much more than just the views of businesspeople. In the latest 2005 update of our governance indicators, our data sources include four cross-country surveys of firms, as well as seven commercial risk-rating agencies, which one might think reflect narrower business interests. But we also rely on three cross-country surveys of individuals, six sets of ratings produced by government and multilateral organizations (such as the World Bank, the African Development Bank, and the U.S. State Department), and finally another 11 data sources produced by a wide range of nongovernmental organizations (such as Freedom House, Reporters Without Borders, and many others). It is therefore simply incorrect to dismiss our indicators as reflecting solely the narrow interests of the business elite.

While we accept that antiregulation biases in surveys of businesspeople are possible in principle, the more relevant question is whether such biases are practically important. Here KS offer us no empirical evidence. This is unfortunate because the various hypotheses of bias that they advance lend themselves well to empirical testing. Consider for example the argument that businesspeople have a view of what constitutes good governance that is fundamentally different from other types of respondents. If this is true, then the responses of firms (or commercial risk-rating agencies who serve mostly business clients) to questions about governance should not be very correlated with ratings provided respondents who are more likely to sympathize with the common good, such as individuals, NGOs, or public sector organizations.

Table 1 Government Effectiveness: Comparing View of Different Respondents

Table reports the correlation of the 2005 assessment of Government Effectiveness from the Global Competitiveness Report's survey of firms with other sources

Surveys of Firms	
World Competitiveness Yearbook	.74
BEEPS	.43
Commercial Risk Rating Agencies	
Economist Intelligence Unit	.86
Political Risk Services	.75
Global Insight DRI	.76
World Markets Online	.88
Merchant International Group	.73
Business Environment Risk Intelligence	.87
Public Sector Agencies	
World Bank CPIA Ratings	.74
African Development Bank CPIA Ratings	.75
Asian Development Bank CPIA Ratings	.42
NGOs	
Columbia University State Capacity Project*	.79
Freedom House	.76
Global E-Government	.53
Bertelsmann Transformation Index	.56
Surveys of Individuals	
Latinobarometer	.52
Afrobarometer	.70

*Data refers to 2004.

This turns out not to be the case. Table 1 reports some simple correlations of assessments of Government Effectiveness from one of our cross-country surveys of firms, the Global Competitiveness Survey (GCS), with all of our other data sources for this measure in 2005.[3] We first consider two other surveys of firms, the World Competitiveness Yearbook and the Business Environment and Enterprise Performance Survey (BEEPS). If businesspeople have a monolithic view of what constitutes good governance, we should expect these two surveys of firms to be highly correlated with the World Competitiveness Yearbook. This is not obviously the case, with correlations of .74 and

[2]While not one of our main points, we find KS's discussion of the possibility that firms' perceptions of onerous and excessively bureaucratic regulation simply reflect their disgruntlement with a benevolent regulatory regime to be a bit naive. See for example Shleifer and Vishny (1999) for an extensive account of "grabbing hand" bureaucrats who use regulation to extract rents for themselves.

[3]We use the individual or average of individual questions from each of listed sources as it enters into our aggregate governance indicators. Details on these specific measures can be found in Appendices A and B of Kaufmann, Kraay, and Mastruzzi (2006b).

.43, respectively. In fact it is notable that the correlation with the BEEPS is lowest among all sources reported in this table. While it is true that the firm survey responses tend to be fairly highly correlated with commercial risk-rating agencies, they are also quite highly correlated with other data sources. It is striking for example that the correlation between firm responses and individual responses from the Afrobarometer survey is .7 across the 23 countries in Africa covered by these two surveys. It is also striking that the correlation of the GCS with the World Bank and African Development Bank's assessments of policy and institutional quality is around .75. While it is true that there are some data sources that are not very highly correlated with the GCS, we do not think one could reasonably conclude from Table 1 that there is a substantial bias in the responses of businesspeople relative to those of other types of respondents. More systematically, in Kaufmann, Kraay, and Mastruzzi (2006b) we document how the country rankings provided by our aggregate indicators are very robust to alternative weighting schemes. This can only be the case if on average there is substantial consensus among our different data sources regarding the broad concepts of governance being measured.

KS also argue that cross-country surveys of both firms and individuals are affected by cultural biases. For example, respondents in different countries might have different norms as to what does or does not constitute corruption. Presumably, however, these cultural biases should not be present in cross-country expert assessments that are deliberately designed to be comparable across countries. This in turn suggests that low correlations between surveys and expert assessments would be consistent with cultural biases. Yet what is striking is that surveys of firms tend to be quite highly correlated with expert assessments of all types. This can be seen in Table 6 of Kaufmann, Kraay, and Mastruzzi (2006b) where we document sizeable correlations between expert assessments and the GCS, for all six of our dimensions of governance.

We have also in our past work looked at other potential sources of bias in our data sources that are not mentioned by KS. One concern often heard is that the ratings provided by NGOs and think-tanks tend to be colored by the ideological orientation of the organization providing the ratings. In Kaufmann, Kraay, and Mastruzzi (2004) we devised a simple test for such political biases. We examined whether the difference between the assessments of think-tanks and firm surveys was systematically correlated with the political orientation of the government in power in the countries being rated. We found that this was generally not the case, casting doubt on this additional possible source of bias.

Even so, we would not want to argue that either cross-country surveys or expert assessments of various sorts are perfect measures of governance. We have long acknowledged that every one of our underlying data sources—or any other potential measure for that matter—is at best an imperfect proxy for governance. This is precisely why we think it is useful to construct aggregate governance indicators that combine information from many different sources: In doing so we are able to smooth out some of this measurement error, with the result that the aggregate indicators are more reliable measures of governance than any of our individual indicators.

Growth and Perceptions of Governance: How Shiny is Your Halo?

KS next argue that perceptions-based measures of governance respond to recent economic performance and thus do not reliably capture deeper notions of institutional quality. This concern, which is sometimes referred to as the "halo effect," is also not new and has been applied to various indicators of governance.[4] KS go on to provide some empirical work which they claim provides evidence of this sort of bias. They estimate cross-country regressions of our Government Effectiveness measure on per capita GDP, a measure of human capital, the logarithm of country population, and per capita GDP growth in the two previous years. The first three variables are intended as controls for fundamental determinants of governance, and the growth variable is intended to pick up "halo" effects. They find that prior growth is significantly correlated with better governance ratings. Based on this they argue that governance indicators based on perceptions data are unreliable because they reflect recent short-run economic performance rather than more fundamental cross-country differences in institutional quality.

We do not find this exercise compelling for several reasons. First, the finding is simply not robust to very minor but reasonable changes in their empirical speci-

[4]See for example Glaeser et al. (2004) who assert without evidence that the correlation between per capita incomes and perceptions-based measures of governance is due to such halo effects. In Kaufmann, Kraay, and Mastruzzi (2006a) we show that this argument is unlikely to matter much empirically because it would require implausibly large halo effects.

TABLE 2 Non-Robustness of KS Halo Effects Regressions

Table reports the coefficient on two-year average prior growth in a regression of the indicated aggregate governance measure on prior growth, per capita income, population, and schooling. The first panel replicates KS Table 1. The second panel replaces per capita income with log per capita income, and a stock of human capital measure with a flow measure of enrollment rates.

	Replication of KS Specification					
	VA	**PV**	**GE**	**RQ**	**RL**	**CC**
1996 Coef	1.28	4.92	2.57	3.21	.07	3.49
t-stat	.89	2.81	2.30	2.27	.06	2.54
1998 Coef	3.31	7.70	3.58	6.56	3.63	3.05
t-stat	1.78	3.98	2.53	4.01	3.05	2.47
2000 Coef	5.70	7.31	5.43	5.46	3.41	3.67
t-stat	2.74	3.42	3.42	2.98	2.59	2.33
2002 Coef	4.77	7.64	5.83	5.32	5.32	3.04
t-stat	2.25	2.86	3.42	3.26	3.17	1.73
2003 Coef	1.44	6.33	3.72	3.58	4.21	2.01
t-stat	.66	2.77	2.34	2.18	2.68	1.27
2004 Coef	.09	5.92	4.25	5.73	5.11	2.61
t-stat	.04	2.83	2.82	3.73	3.45	1.70
2005 Coef	−.23	5.24	4.25	5.77	4.62	2.83
t-stat	−.10	2.47	2.83	3.76	3.10	1.86
# Significantly > 0:32						
# Significantly < 0:0						
	Results for 150 Country Sample					
	VA	**PV**	**GE**	**RQ**	**RL**	**CC**
1996 Coef	1.94	3.46	2.23	2.50	1.07	2.52
t-stat	1.91	3.17	2.48	2.86	1.24	2.46
1998 Coef	−1.41	5.03	.05	−.55	−1.25	−.22
t-stat	−1.22	3.26	.05	−.52	−1.34	−.22
2000 Coef	−1.25	2.87	−2.41	−1.43	−2.59	−2.85
t-stat	−1.18	1.79	−2.65	−1.38	−2.99	−2.93
2002 Coef	−1.01	3.84	.69	1.34	.12	−1.34
t-stat	−.67	2.18	.59	1.07	.09	−.97
2003 Coef	−3.09	2.52	.39	.76	−.78	−2.44
t-stat	−1.93	1.57	.32	.59	−.63	−1.85
2004 Coef	−3.49	.43	−1.85	−.19	−2.33	−4.72
t-stat	−2.19	.31	−1.59	−.15	−1.91	−3.67
2005 Coef	−3.72	.49	−1.25	.40	−2.37	−3.97
t-stat	−2.28	.34	−1.06	.33	−1.92	−3.03
# Significantly > 0:6						
# Significantly < 0:7						

fication. We show this in Table 2. In the first panel we summarize a set of regressions that essentially replicate KS's Table 1. In particular we replicate the first four cross-sectional regressions from that Table, with Government Effectiveness as the dependent variable.[5] We estimate these regressions for 1996, 1998, 2000, and 2002 as do KS, and also for 2003, 2004, and 2005 using the most recently available update of our governance indicators. Moreover, we think that KS's halo effects critique could in principle apply to all six of our aggregate governance indicators, so we also estimate

[5]Our results here are not exactly the same as those reported by KS because of minor discrepancies between our data sets. We use a slightly revised newer version of our governance indicators and also a slightly different measure of human capital from the Barro-Lee data set. However, our results mirror exactly the pattern of significance that KS show and so we think constitute a "fair" replication of their finding.

the same regressions for our other five indicators. This results in a set of 42 regressions. To conserve space we report only the coefficient on lagged growth, so that each cell in the table refers to a separate regression. Consistent with KS we find that for our Government Effectiveness measure, lagged growth is significantly positively correlated with the governance indicator (third column of top panel of Table 2). This is true not only for the years 1996–2002 considered by KS, but also for 2003–2005. Moreover, the other columns show that in many cases, lagged growth is significantly correlated with our other governance indicators: In 32 of the 42 regressions we find a significant positive correlation at the 5% significance level.

In the next panel we show the effect of two minor departures from the original KS specification. Instead of entering per capita GDP in levels as they do, we enter it in log-levels. This is very standard practice in cross-country empirics and statistically is more appropriate since the relationship between the dependent variable and log per capita GDP is much closer to being linear, and we are using a linear regression model. In addition, we note that KS's results cover a relatively small set of around 100 countries. The key constraint here is the limited availability of the human capital variable that they use—the governance indicators are available for over 200 countries. We therefore also expand the sample of countries by replacing the stock of human capital variable with a more widely available flow measure: the gross secondary school enrollment rate.[6]

The new regressions in this panel cover a much larger sample of around 150 countries. Now the results are drastically different from before, in that we find only six of 42 cases where lagged growth enters significantly and positively. And more peculiarly, there are seven cases where lagged growth enters significantly *negatively*, which is just the opposite of what one would expect if halo effects are important. In the remaining 29 cases there is no significant correlation between our governance measures and growth in the previous two years. In fact, looking only at the sign of the estimated coefficients, we find that they are nearly evenly split between 24 positive and 18 negative. The fact that these two minor—but we think very reasonable—variants eliminate the significance of KS's results suggests to us that their findings are simply not robust. Or put differently, in order for KS's critique to be convincing they would need to also provide an account of why halo effects are important in their particular 100-country sample but not in a broader 150-country sample.

Nevertheless, suppose we take the KS sample and specification at face value, despite its lack of robustness. We next argue that KS are mistaken to interpret the significance of the coefficient on lagged growth as evidence of halo effects, because we might very well observe this partial correlation even if halo effects were not present in the data. The reason is simple. Suppose, reasonably enough, that countries with effective governments do grow faster, at least in the long run. Suppose, also reasonably, that government effectiveness is fairly persistent over time. Then we should expect to see a correlation between long-run growth, for example, growth over the previous 10 or 20 years, and our measures of Government Effectiveness. This would not be due to halo effects, but rather would simply reflect the beneficial growth effects of having a competent government over a long period of time. However, to the extent that recent growth performance is correlated with long-run growth performance, we might very well find that the former is in fact correlated with our measure of Government Effectiveness simply because the latter is omitted from the regression.

This omitted variable problem turns out to be important in KS's preferred specification. The easiest way to see this is to again replicate the regressions in KS's Table 1, but now adding a variable capturing long-run growth in the 20 years prior to the date of the governance indicator. We do this in Table 3, for KS's basic specification for Government Effectiveness. In all of the five periods shown, we find that prior 20-year average growth is significantly correlated with Government Effectiveness. Moreover, now in only one case is recent growth performance significantly correlated with government effectiveness.

One could nevertheless still argue that the significance of prior long-run growth in these regressions constitutes evidence of halo effects, with subjective perceptions of government effectiveness being tainted by previous long-run economic performance as well. However, we do not think this would be credible. To show this, we also estimate the same regression, but instead replace the dependent variable with the Evans and Rauch (1999) measure of professionalism in the civil service, which KS laud as a carefully constructed measure of institutional quality likely to be free of any perceptions biases (although unfortunately covering only a small sample of 35 countries at one point in

[6]Although this flow measure is of course conceptually different from the stock of human capital, the two are quite highly correlated across countries. And since KS do not offer a theoretical model which insists that the stock measure be used, we think it is reasonable also to look at the flow.

TABLE 3 Nonrobustness of KS Halo Effects Regressions, Cont'd

This table replicates the regressions in KS Table 1 but adds prior 20-year average growth as a right-hand-side variable.

Dependent Variable	GE 1996	GE 1998	GE 2000	GE 2002	GE 2004	Evans-Rauch
Per Capita GDP	.096	.087	.074	.068	.073	.240
	(.010)***	(.010)***	(.009)***	(.008)***	(.007)***	(.106)**
Human Capital	.038	.032	.033	.072	.052	−.190
	(.031)	(.034)	(.031)	(.028)**	(.027)*	(.328)
2-year prior growth	.689	1.935	3.234	1.668	1.453	−6.100
	(1.383)	(1.624)	(1.618)**	(1.946)	(1.643)	(18.936)
20-year prior growth	5.851	5.818	9.934	1.228	9.236	67.001
	(2.612)**	(3.008)*	(2.609)***	(2.655)***	(2.643)***	(27.238)**
log(population)	−.009	−.054	−.011	−.033	−.035	.805
	(.029)	(.032)*	(.030)	(.026)	(.026)	(.385)**
Constant	−.765	.006	−.692	−.508	−.440	−8.609
	(.485)	(.540)	(.507)	(.441)	(.436)	(6.916)
Observations	93	94	97	98	97	31
R-squared	.86	.82	.83	.87	.87	.53

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%.

time). Again we find that prior long-run growth performance is significantly correlated with this measure of institutional quality which is unlikely to be tainted by any kind of halo effects. And once we control for long-run growth, short-run growth has no significant correlation with this measure of institutional quality. This supports our interpretation of the data: Government effectiveness is both persistent over time and correlated with long-run growth, which in turn is correlated with short-run growth. This explains the spurious significance of short-run growth in the KS regressions.

Suppose despite all this one were to insist on interpreting the significance of short-run growth in KS's specifications as evidence of halo effects. A final question one might ask is whether these effects are practically important or not, in the sense of significantly contributing to the variation in our governance indicators. We answer this question by calculating the share of the cross-country variation in our Government Effectiveness measure that is due to the estimated halo effect, defined as prior growth multiplied by its estimated coefficient.[7] We find that this share is quite small, ranging from a low of 1.6% in 2004 to a high of 5.6% in 2000. While we have argued at length that the KS's results are not robust and their interpretation is flawed, even if we take them at face value, the halo effects they claim to document do not appear to be quantitatively large.

Governance and Growth: Putting the Straw Man Out of His Misery

In the final part of their paper KS investigate the relationship between our measure of Government Effectiveness and subsequent growth, in a large cross-section of countries. The left-hand-side variable in their regression is real per capita GDP growth in the two years following the date of the governance indicator. The right-hand-side variables consist of our Government Effectiveness measure and a set of other control variables. They make much of the fact that, conditional on these other variables, good scores on the Goverment Effectiveness indicator are not correlated with faster subsequent growth. In fact, they conclude from this that there is no evidence that good governance raises growth performance in the next two years.

We think this final exercise is little more than a straw man, for several reasons. To begin, we note that KS appear to completely bypass a large and careful literature that has studied the connection between institutional quality and long-run economic performance. For example, in a seminal empirical paper Knack and Keefer (1995) documented a highly significant partial correlation between various measures of

[7]The variance share is defined as (VAR (halo effect) + COV (halo effect, government effectiveness))/VAR (government effectiveness). See Klenow and Rodriguez-Clare (1997) for a justification.

institutional quality and 25-year average growth rates across countries, controlling for a variety of factors. In another highly influential paper, Acemoglu, Johnson, and Robinson (2001) have shown that the historically determined component of institutional quality has had a strong causal effect on current levels of per capita income across countries today. Since cross-country differences in per capita incomes today primarily reflect differences in these countries' very long-run growth rates, this paper can also be interpreted as capturing a relationship between institutional quality and subsequent very long-run growth. These two papers, as well as many others in this literature, have helped to shape the views of many academics and policymakers that good governance is important for growth.[8]

Should this view be shaken by the couple of regressions that KS offer? We think not. First, we note that KS's choice of very short-run growth as the dependent variable differs from most of the existing literature which has focused on long-run growth. We think this latter focus is much more appropriate. It seems to us quite plausible that the growth effects of good institutions show up only over time, while short-run fluctuations unrelated to institutional quality are likely to dominate a lot of the variation in yearly growth rates. It is for this reason that virtually the entire empirical cross-country growth literature has focused on long-run average growth rates. In contrast, KS offer no justification for looking at the relationship between very noisy short-run fluctuations and governance.

Second, we have already noted that the literature on institutions and growth has—for good reasons—mostly focused on more fundamental notions of protection of private property as a proxy for good governance. For example, the Acemoglu, Johnson, and Robinson (2001) paper emphasizes the importance of property rights protection as proxied by a measure of expropriation risk. We recognize that various measures of property rights protection tend to be fairly correlated with our Government Effectiveness measure that KS use, and it is likely that KS would obtain similar results in their problematic specification if they used these other measures. However, we think that KS could do a much better job of justifying their focus on just this one very particular measure of governance to the exclusion of arguably more fundamental ones studied by the existing literature.

[8]See for example Hall and Jones (1999) and Rodrik, Subramanian, and Trebbi (2004) for other contributions to this literature.

Third, KS depart from most of the recent literature on institutions and growth by sidestepping entirely the fundamental question of causality—do observed correlations between governance and growth reflect causation from governance to growth or the other way around?[9] Or do they reflect the effect of a myriad of potential other variables not included in the regression that drive both growth and governance? This econometric difficulty is by now very well understood, and leading papers in the literature have come up with a variety of creative strategies for sorting out the causal effect of good governance.[10] The strategy of naively estimating cross-country regressions by ordinary least squares alone as is done by KS has long been abandoned by the serious empirical growth literature.[11]

Finally, we note that the regressions presented by KS contain some quite implausible estimates of the effects of other growth determinants. Should we take seriously their findings on governance and growth when the very same regressions also tell us that macroeconomic stability (as captured by inflation) is insignificantly correlated with growth, whereas the empirical growth literature has showed that at least very high inflation rates are correlated with slower growth (columns 1–4 of their Table 3)? Should we take seriously their finding in column 4 of Table 3 that higher investment rates are significantly *negatively* correlated with growth? If KS would like us to take seriously the lack of significance of governance in their

[9]This omission is particularly troubling given that they devote the entire middle part of their paper to arguing that there is in fact one such channel of reverse causation, through halo effects.

[10]Leading examples are the construction of creative instrumental variables, such as in Acemoglu, Johnson, and Robinson (2001), the use of time-series exclusion restrictions such as in Chong and Calderón (2000), and the use of identification through heteroskedasticity as in Rigobon and Rodrik (2004).

[11]KS also ignore the difficulty of estimating growth regressions, which are intrinsically dynamic panel regressions, using ordinary least squares. As is well understood, in the presence of unobserved country effects, initial income is by construction correlated with the error term, and this endogeneity problem in general contaminates estimates of all of the coefficients of interest. See Caselli, Esquivel, and Lefort (1996) for an application to growth empirics. We also find KS's choice of estimation technique for their pooled regressions peculiar. In the last column of their Table 1 they choose a random effects estimator, while in the last column of their Tables 3 and 4 they choose a fixed-effects estimator. As noted above neither a fixed-effects nor a random-effects estimator will yield consistent estimates of the slope coefficients in the growth regressions in Tables 3 and 4. However, we find it interesting that had KS used the same random effects estimator in Table 3 as they had in Table 1, they would have found that government effectiveness *actually enters positively and significantly for growth in their own specification*.

growth regression, they should also provide an account for these other quite peculiar findings.

In conclusion, we do not wish to argue that the empirical literature on governance and growth is now conclusive so that no further work is needed—to do so would be complacent in the extreme. We are acutely aware of the limitations of existing cross-country measures of governance and have long argued for the need to complement cross-country indicators with more detailed and nuanced within-country data in order to inform efforts to improve governance at the country level. There is also plenty of room for further serious work in understanding both the causes and consequences of good governance, at the cross-country and within-country level. We do however think that contributions to this exciting and important research agenda would do better to take the large existing literature as a starting point and then document whether these findings can be overturned—or refined—in reasonable ways. Starting *de novo* with a few flawed growth regressions as KS do seems to us unlikely to be helpful in advancing our understanding of these important issues.

Acknowledgment

The opinions expressed here are the authors' and do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

Manuscript submitted 22 August 2006
Manuscript accepted for publication 29 August 2006

References

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development." *American Economic Review* 91 (5): 1369–1401.

Caselli, Francesco, Gerardo Esquivel, and Fernando Lefort (1996). "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics." *Journal of Economic Growth* 1 (1): 363–89.

Chong, Alberto, and Cesar Calderón. 2000. "Causality and Feedback Between Institutional Measures and Economic Growth." *Economics and Politics* 12 (1): 69–81.

Evans, Peter, and James Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review* 64: 748–65.

Glaeser, Edward, Rafael LaPorta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271–303.

Hall, Robert E., and Charles Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (1): 83–116.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999a. "Aggregating Governance Indicators." *World Bank Policy Research Working Paper No. 2195*. Accessible at http://www.worldbank.org/wbi/governance/pubs/aggindicators.html.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999b. "Governance Matters." *World Bank Policy Research Working Paper No. 2196*. http://www.worldbank.org/wbi/governance/pubs/govmatters.html.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 2002. "Governance Matters II—Updated Indicators for 2000/01." *World Bank Policy Research Working Paper No. 2772*. http://www.worldbank.org/wbi/governance/pubs/govmatters3.html.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2004. "Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002." *World Bank Economic Review* 18 (2): 253–87.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2006a. "Measuring Governance Using Perceptions Data." In *The Handbook of Economic Corruption*, ed. Susan Rose-Ackerman. Northampton, England: Edward Elgar.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2006b. "Governance Matters V: Governance Indicators for 1996–2005." *World Bank Policy Research Department Working Paper No. 4012*.

Klenow, Peter, and Andres Rodriguez-Clare. 1997. "The Neoclassical Revolution in Growth Economics: Has it Gone Too Far?" In *NBER Macroeconomics Annual*, eds. Ben Bernanke and Julio Rotemberg. Cambridge: MIT Press, pp. 73–103.

Knack, Steven, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-Country Tests Using Alternative Measures." *Economics and Politics* 7 (3): 207–27.

Rigobon, Roberto, and Dani Rodrik. 2004. "Rule of Law, Democracy, Openness, and Income: Estimating the Interrelationships." Typescript. Massachusetts Institute of Technology.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9 (2): 131–65.

Shleifer, Andrei, and Robert Vishny. 1999. *The Grabbing Hand: Government Pathologies and their Cures*. Cambridge: Harvard University Press.

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi work at the World Bank Group, Washington, DC 20433.

Growth and Governance: A Defense

Marcus J. Kurtz The Ohio State University
Andrew Schrank University of New Mexico

It is clear we have disagreements from our esteemed colleagues on this important issue. The Kaufmann, Kraay, and Mastruzzi [hereafter KKM] response raises a number of insightful points that advance this debate. We thank them for that and for taking the time to engage our article. As we have reflected on our original essay and KKM's critique of it, we were struck by the fact that we may actually be talking past each other and that differences in approach between political science and economics may foster this miscommunication. We are not arguing that political science offers a superior approach. To the contrary, our intellectual debt to economics is enormous. Our bibliography offers testimony to that very fact. We nevertheless believe that political scientists have as much to offer economists as vice versa, including not only a venerable literature on state formation and governance but a no less important tradition of self-conscious concept formation that has been less central to most economists.[1] While political scientists have traditionally been producers of political data, and have therefore treated the process of concept formation as a necessary prelude to both measurement and modeling (Collier and Mahon 1993; Gerring 1999; Sartori 1970), economists have until recently been consumers of political data and have, therefore, subordinated the need for self-conscious concept formation to the understandable urge to measure and model. We ultimately hope to demonstrate that conceptual issues—and corresponding measurement problems—are at the core of the debate over growth and governance and that progress will be less than ideal until they are addressed. A meeting of minds may be too much to ask for at the present time, but a meeting of method would almost certainly constitute a step in the right direction. We, therefore, offer this rebuttal as not just a response to the thoughtful critique of KKM, but as an effort to bridge these gaps and develop even better ways to make real advances in understanding the relationship between growth and governance.

Concept Formation

KKM defined government effectiveness (GE) as the "competence of the bureaucracy and the quality of public service delivery" in their earlier papers (KKM 2005, 4). Our doubts about GE derived neither from their definition nor from their conceptual scheme but from the construction of the measure itself. We worried that perception-based indicators of governance in general, and their indicator of GE in particular, suffered from systematic measurement error, selection bias, and halo effects—in short that they did not measure the concept they introduced.

By way of rebuttal, however, KKM have all but abandoned their indicator of GE in favor of distinct indicators (e.g., Rule of Law) designed to capture "a more basic notion of governance going back to the seminal work of Douglas North: the norms of limited government that protect private property from predation by the state" (KKM 2007, 553). We therefore find ourselves in the awkward position of defending their conceptualization of GE in order to criticize their operationalization of GE. Why is an ostensibly narrow focus on the "competence of the bureaucracy" superior to an allegedly encompassing focus on the "protection or private property" (KKM 2007, 559)? We believe that political institutions (e.g., bureaucracy) are conceptually independent of policy decisions (e.g.,

[1]An admittedly unscientific review of graduate methodology syllabi in the top five political science and economics departments confirms our suspicions. Political science departments almost invariably include at least one course with a substantial component on concept formation and measurement. Economics departments no less consistently focus on econometrics and modeling. For a theoretical analysis of important political science concepts, see, e.g., Ball, Farr, and Hanson (1989).

The Journal of Politics, Vol. 69, No. 2, May 2007, pp. 563–569

 ISSN 0022-3816

expropriation) and that the former should be defined and evaluated without regard to the latter. A government that ably protects property, lowers taxes, or privatizes industry is not necessarily more capable than a government that adroitly expropriates property, taxes income, or nationalizes industry—whatever one thinks of the policies in question.

We are by no means the first social scientists to distinguish the enduring rules, norms, or constraints implied by the word "institution" from the at times transitory decisions of powerful public officials. After all, Edward Glaeser and his collaborators have already portrayed estimates of "expropriation risk" derived from commercial data sources employed by KKM as inadequate indicators of political institutions. "Whatever expropriation risk measures," they argue, "it is obviously not permanent rules, procedures, or norms supplying checks and balances on the sovereign" (Glaeser et al. 2004, 276).

Our point is neither to provoke nor to belabor a "terminological tussle" (KKM 2007, 553) but to acknowledge and underscore the fact that "concept formation stands prior to quantification" (Sartori 1970, 1038). While KKM take comfort in the fact that their various indicators are highly correlated with each other and therefore portray the "protection of private property as a proxy for good governance" (KKM 2007, 559)—if nothing else—in their rebuttal, their optimism is arguably misplaced, for bivariate correlations speak to the reliability, rather than the validity, of political indicators (see, e.g., Munck and Verkuilen 2002, 29), and their preferred measure of property protection (i.e., their Rule of Law indicator) is triply problematic in any event. First, they exaggerate the transparency and intelligibility of property claims and thereby add an additional source of error to their indicators. After all, the definition of property is at best controversial and at worst socioculturally bounded. A government that evicts squatters will in all likelihood be portrayed as a threat to private property by the squatters and a bulwark against expropriation by the landlords—and the problem is likely to be compounded by the fact that in much of the developing world this year's squatters are likely to be next year's landlords and vice versa. Second, they conflate short-run policy preferences like the protection of private property with enduring institutions like meritocratic bureaucracies and thereby render the assessment of causal effects all but impossible. Are the significant coefficients obtained by KKM—and consumers of their data—attributable to the narrowly institutional features of their preferred indicators or to simultaneously incorporated, and therefore statistically indistinguishable, policy considerations? Unfortunately, the answer is anything but clear. And, finally, they incorporate policy outcomes (as well as policy preferences) across a wide range of different issue areas and thereby aggravate the already vexing problem of causal inference. In fact, KKM define the Rule of Law as "the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police and the courts, as well as the likelihood of crime and violence" (KKM 2006, 4). Positive RL coefficients are therefore no more interpretable than they are surprising. Do they reflect the impartiality of the police and courts? The benefits of crime control? The returns to social capital and generalized trust? The fact that crime and violence are associated with a host of related social ills? Or the costs of expropriation, tax evasion, informality, patent and copyright infringement, or any of the other transgressions specified in the actual measure? And how would one know?

The literature on concept formation in comparative politics holds that effective concepts discriminate among closely related behaviors, outcomes, or processes and that "for fact-finding purposes it is more profitable to exaggerate in over-differentiation than in over-assimilation" (Sartori 1970, 1039). But KKM take an ambivalent stance toward discrimination. On the one hand, they disaggregate "governance" into six different dimensions and, to their inestimable credit, warn consumers of their data not to aggregate their various indicators into an overall index of "good government." On the other hand, they incorporate institutions, policy preferences, and policy outcomes into the definitions of their preferred indicators and thereby undermine the very benefits of disaggregation.

Furthermore, KKM are growing *less* discriminating over time. Take, for example, GE. While KKM originally defined government effectiveness by way of reference to relatively coherent aspects of bureaucratic organization and behavior, and thereby bracketed the behavior of the legislature, the sovereign, and the courts, they have recently added "the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (KKM 2006, 4) to their definition, and have thereby brought at least two of the aforementioned three actors, as well as public policies and subjective assessments of their "credibility," back into the picture. Nevertheless, GE continues to constitute the closest thing KKM have to the sort of purely institutional measure that political scientists routinely investigate and it therefore offers a "best case" scenario for the KKM indicators more generally.

Measurement and Sampling

We maintain that perception-based indicators of governance in general, and the KKM indicator of GE in particular, are overly dependent on the impressions of businesspeople. KKM offer a twofold rebuttal. First, they maintain that they incorporate data not only from businesspeople but from a broader array of sources (e.g., nongovernmental organizations, citizen surveys, etc.) into their indicators. And, second, they question the degree to which business and societal preferences diverge in any event by noting the high correlation of business and social surveys.

The former claim is puzzling for several reasons. First, they cite a number of sources (e.g., Reporters without Borders, the U.S. State Department) that are incorporated *not* into their most recent indicator of GE but into distinct measures that are not in dispute. Second, they fail to distinguish "representative" sources that are available for the majority of their countries from "nonrepresentative" sources that are not—and that perforce carry less weight in their overall indicators. Only one of seven so-called representative sources in the most recent release of GE is an NGO, for example, and none is a citizen survey (KKM 2006, Table B3). Six of the seven more widely available sources are commercial risk-rating agencies or surveys of businesspeople that betray the myriad biases we have already described and therefore validate—rather than assuage—our original fears. Third, they exaggerate the neutrality of their noncommercial sources in any event. After all, the Freedom House ratings they use have been portrayed as an "ideologically loaded" product of American values and preferences by Adam Przeworski (Munck n.d., 38; see also Munck and Verkuilen 2002). The multilateral development banks are by no means policy neutral (Wade 2002a). And indicators of "electronic governance" assume—rather than prove—that e-governance is good governance and are biased against countries on the other side of the "digital divide" by construction (Wade 2002b). Finally, they fail to acknowledge the full implications of their aggregation scheme. On the contrary, KKM explicitly *assume*—but fail to prove—that the errors of their distinct sources are independent of each other and therefore employ an aggregation rule that explicitly "rewards conformity" (Kaufmann, Kraay, and Zoido-Lobatón 1999, 14). Sources that deviate from common patterns are given less weight in their aggregate indicators than sources that are highly correlated with each other, and the two citizen surveys that weigh so heavily in their response—Afrobarometer and Latinobarometer—are therefore responsible for about 2% of their overall rating of GE (KKM 2003, Table 3).[2] While KKM find that their weighted measure of GE and an "equally weighted" alternative are correlated at $r \geq .97$, and thereby conclude that their weights are immaterial to their estimates (KKM 2006, 25), their alternative measure remains disproportionately dependent upon the perceptions of businesspeople and their advisers as well. After all, GE includes a single citizen survey from each of 36 countries—18 in Latin America and 18 in Africa—and no citizen surveys from Asia, Europe, North America, or Oceania. Consequently, the equally weighted alternative *excludes* the voices of citizens in well over 75% of the countries of the world—including eight of the ten largest countries—and relies *disproportionately* upon commercial sources for the 36 countries whose rankings *do* include citizen surveys, since all 36 incorporate ratings from more than one commercial source. Businesspeople and their advisors constitute a tiny—and by most accounts politically distinct—minority of the world's population but contribute a vastly disproportionate share of the GE index under *either* weighting scheme, and we are therefore neither surprised nor convinced by the high reported correlation—which is a test of reliability rather than validity in any event.

Of course, KKM's biased sample is immaterial if errors are really independent and perceptions of GE are impervious to the respondent's occupation or position. What, then, are we to make of their Table 1 (KKM 2007, 554)? Do the reported correlations prove that firms and citizens perceive their public sectors in the same way? We preface our answers to both question with a by now redundant caveat—that bivariate correlations are in and of themselves tests of reliability rather than validity—and a benchmark drawn from the expansive literature on the conceptualization and measurement of democracy. The various measures of democracy designed and used by political scientists are correlated at $r \geq .8$ (Przeworski et al. 2000, 56–57). While the Afrobarometer survey and the GCS data invoked by KKM betray a reasonably high correlation ($r = .7$), the former is deliberately weighted toward 18 atypically liberal African regimes, and the more comprehensive Latin American data that go all but unremarked upon in their text diverge markedly. In fact, the correlation between the Latinobarometer survey of citizens *throughout* noncommunist Latin America and the GCS survey of firms—allegedly measures of the *same* concept—is a modest .52, and the mean correla-

[2]KKM released weights that would be applied "to a hypothetical country appearing in all of the available sources for that indicator" in 2002 (KKM 2003, 45).

TABLE 1 Do Businesspeople and their Neighbors Perceive Government in the Same Way?

Question	Responses	Businessperson
Based on your experiences, how easy or hard is it to obtain household services (like electricity or telephone)?	0 = never try, very difficult, or difficult; 1 = easy or very easy	1.755 ($p = .000$)
Based on your experiences, how easy or hard is it to obtain identity documents (like birth certificate, passport)?		1.227 ($p = .001$)
How well or badly would you say the current government is handling health care?	0 = very badly or fairly badly; 1 = fairly well or very well	.889 ($p = .053$)
How well or badly would you say the current government is handling education?		.777 ($p = .000$)

Self-identified businesspeople are coded 1; others are coded 0; country dummies are suppressed; and odds ratios for businesspeople are presented next to their parenthesized p values. We have dichotomized the 4- and 5-point Likert scales of answers for ease of calculation and interpretation and dropped respondents who were unwilling or unable to answer. The data are from Round 2 of Afrobarometer (2002–2003); the more recent data available to KKM are not publicly available.

tion between the Latin American data and the six commercial risk assessments found in their Table 1 is a mere .42.[3]

Nor is Latinobarometer exceptional. The mean correlation for the eight commercial risk assessments and firm surveys incorporated into their Table 1 is .75. The mean correlation for the eight noncommercial sources (e.g., public agencies, NGOs, citizen surveys) that are both included in the table and incorporated into their most recent GE indicator is .62. And the difference in means is significant at $p \leq .08$.[4]

Nevertheless, the Afrobarometer data permit a simpler test of KKM's faith in the insignificance of the differences between business and popular perceptions of GE. Afrobarometer includes data on whether the respondent is a "business person." Table 1 (above) includes logistic regressions of individual responses to all of the Afrobarometer questions included in the most recent GE indicator on an indicator variable coded 1 if the respondent is a business person and a series of country dummies. Positive assessments are coded 1; businesspeople are coded 1; odds ratios are placed next to their parenthesized *p* values; and the invariably significant results suggest—contra KKM—that businesspeople have *better access* to government services and nonetheless hold their governments in *lower regard* than their compatriots. What are the implications for GE? The fact that businesspeople and citizens within the same country part company on the question of good governance speaks to a deeper issue that underpins the entire debate: We don't even have common perceptions of government effectiveness at the national level let alone the ability to construct cross-nationally valid rankings.[5]

Furthermore, the Afrobarometer data present a relatively conservative test of our claim. After all, the survey questions cover more or less uncontroversial issues like household services, education, and health care provision. The risk-rating authorities who are responsible for the bulk of the GE indicator address more controversial topics like taxation, regulation, and the existence and interpretation of red tape—where the interests of firms and citizens are particularly likely to diverge.

In fact, KKM's colleagues at the World Bank devote an entire subsection of a recent *World Development Report* to the "basic tension" between "firm preferences" for limited taxation, regulation, and spending and the overarching "public interest" (World Bank 2004, 37). Why and to what effect would KKM depart from the position taken by their employer's "flagship publication" (Wade 2002a, 220)? We worry that KKM are implicitly nullifying a decade of *glasnost*

[3]Only two of six correlations between Latinobarometer and commercial risk assessments are significant at $p \leq .05$ and two more at $p \leq .10$. The correlations confirm the reliability but not the validity of the measures in any event. Commercial and noncommercial sources could diverge markedly on average and nonetheless correlate highly.

[4]The inclusion of the Columbia University data would drop the p value to .11; however, the Columbia data have been dropped from the most recent GE indicator and their inclusion in KKM's Table 1 is therefore puzzling.

[5]We couldn't even use the KKM data to construct ordinal rankings without assuming that the direction and degree of perceptual differences between citizens and businesspeople are constant across countries and uniform across indicators. The former is unlikely, and the latter is demonstrably untrue in the Afrobarometer data—as well as unknowable for the bulk of the world's population in light of the paucity of citizen surveys in their data set.

in the international donor community. After all, the multilateral development banks no longer deny industrial policy's potential contribution to growth and development; instead, they hold that industrial policy demands better governance than the typical less-developed country (LDC) can muster and therefore portray free market reform as a second-best alternative. By treating laissez faire as part of the very definition of government effectiveness, however, KKM and their adherents condemn industrial policy by tautology: Where does industrial policy work? Where governments are effective. What defines effective government? The absence of red tape and regulation—that is, the absence of the very lifeblood of industrial policy.

Halo Effects Revisited

We worried that perception-based measures of governance were also contaminated by halo effects. KKM responded by demonstrating that our statistical results were vulnerable to defensible—if by no means unassailable—changes in specification and that GE responds not to "recent growth performance" but to "long-run growth in the 20 years prior to the date of the governance indicator" (KKM 2007, 557). We have neither the space nor the inclination to debate the relative merits or interpretation of different model specifications at length (however see Kurtz and Schrank 2006) and would rather take the opportunity to make two simple but by no means unimportant points about concept formation.

First, we need not win the debate over halo effects to win the debate over GE. We have already established that GE (1) conflates policy preferences and outcomes with political institutions and (2) relies almost entirely upon the perceptions of businesspeople and their advisers in any event. Consequently, the correlation between GE and growth—if any—is at best uninterpretable *whether halo effects are demonstrable or not.* A positive GE coefficient could reflect any combination of at least three different underlying processes: first, a positive relationship between the *policies* captured by the measure (e.g., deregulation, tax relief, school spending, etc.) and growth; second, a positive relationship between *actual government efficacy* (e.g., bureaucratic capability) and growth; or, third, *herd behavior* on the part of investors who receive their advice from the same consultants and risk-rating agencies and thereby animate growth regardless of the so-called fundamentals. We need not win the debate over halo effects, therefore, to win the debate over GE, for conceptualization and measurement are, as political scientists have long noted, logically prior to hypothesis testing. We need only demonstrate that their indicator of GE relies disproportionately upon the perceptions of businesspeople *or* conflates policy preferences and outcomes with political institutions. We do both.

Second, the models presented in KKM's Table 2 (2007, 556)cast at least as much doubt on the validity of GE as they do upon the likelihood of halo effects. Why? KKM depart from past practice, including their own past practice (see, e.g., Dollar and Kraay 2002; Isham and Kaufmann 1999), by abandoning our direct measure of the stock of human resources for a "more widely available" indicator of their flow: the gross secondary enrollment rate. But the gross secondary enrollment rate is more accurately portrayed as a *reflection* than a *cause* of government effectiveness—especially in light of their operationalization of GE. After all, KKM incorporate "the quality of public service delivery" into the very definition of GE (KKM 2005, 4), and their reconstituted models therefore treat a *subjective* measure of service delivery (i.e., GE) as a function of an *objective* measure of service delivery (i.e., enrollment rates). We are not the first social scientists to realize that secondary enrollment constitutes a measure, rather than a proximate cause, of government effectiveness (Migdal 1988, 286), and we are therefore puzzled by the objective measure's persistent inability to predict the subjective measure with confidence.[6] Are the insignificant coefficients on the enrollment variables in the models distilled into their Table 2 products of specification error on the right-hand side or the outright invalidity of the measure on the left-hand size? We fear the former. We suspect the latter. And we note that neither interpretation bodes well for KKM.

The Validity and Virtue of Prospective Models

Our models of prospective growth are designed to mitigate the problem of endogeneity as well as the consequences of halo effects. While we believe that endogeneity is less vexing and halo effects are less likely in prospective models and therefore regress growth rates on prior measures of GE in Tables 2–4 of our original paper, KKM assert "that the growth effects of good institutions show up only over time" and thus call the very basis "for looking at the relationship between very noisy short-run fluctuations and gover-

[6]KKM present only the coefficients on the halo effects. We therefore replicated their models with their data and found that the *minimum p* value for secondary enrollment is .343.

nance" into question (KKM 2007, 559). We are well aware of the noisy nature of short-run growth rates but are nonetheless surprised by KKM's willingness to admit that GE is unlikely to be perceptibly related to prospective short-run growth on average—especially since *their* measure incorporates *several* questions about anticipated *short-run* growth rates (KKM 2006, Table B3) and thereby (1) provides the allegedly absent "justification" for examining prospective growth over short periods and (2) is biased in favor of a *positive* finding from the outset.[7]

Why, then, does their faith in their measure survive the appearance of a *negative* finding? Shouldn't the apparent absence of a correlation between governance and short-run growth call their conviction into question? We believe the results should at least cast doubt upon the consensus and open the door to conceptual and causal reflection. But KKM give the benefit of *any* doubt to students of growth and governance over the *very* long run who *at a minimum*: (1) make heroic assumptions about the worldwide distribution of income circa 1800; (2) conflate the risk of expropriation (i.e., a policy) with the effectiveness of government (i.e., an institution); and (3) instrument for the risk of expropriation in the late twentieth century with the mortality rates of the most violent expropriators in world history—European colonists—in earlier centuries and thereby implicitly illustrate—but fail to take account of—the socioculturally bounded nature of property claims. We've already discussed the limits to the by no means unassailable (see, e.g., Bardhan 2005; Glaeser et al. 2004) existing literature in detail (see, e.g., Kurtz and Schrank 2006, notes 7 and 10) and see no need to revisit the issue other than to note that by invoking a consensus to refute a challenge to the consensus, KKM are pioneering a new and to our minds intolerably conservative approach to social scientific argumentation.

Conclusion

We thank KKM for drawing renewed attention the problem of governance in developing countries and for opening the door to continued and productive interdisciplinary dialogue on the matter. Political sociologists from Max Weber onward have drawn a distinction between the institutional "rules of the game" and the policies produced by the players. While KKM accept the distinction in theory, and portray their estimates as indicators of institutions, they (1) elide the distinction in practice by incorporating policy preferences and outcomes into their indicators and simultaneously (2) compound their error by using a biased sample of sources. The results are triply troubling. First, KKM render the interpretation of their indicator all but impossible. Do positive GE coefficients—if any—reflect bureaucratic capacity, the influence of policy preferences (or outcomes) that are implicitly or explicitly incorporated into their measures (e.g., letting foreign firms "go about their business"), halo effects, or the potentially irrational exuberance of ill-informed investors? And how would one know? Second, KKM send LDC policymakers mixed signals. After all, the MDBs take pride in their widely acknowledged influence over the LDCs (see Kurtz and Schrank 2006, 23), and the United States government is already conditioning aid allocations on performance standards established in part by KKM. The intuition is straightforward: LDCs should be rewarded for good behavior and punished for the opposite. But what is good behavior? What policies will foster an improved GE score? And will they foster growth and development as well? Unfortunately, the answers are anything but clear, for the sources employed by KKM reward aid recipients for policies that are almost certainly in tension with each other including stabilizing their polities, deregulating their markets, lowering their tax rates, ensuring the health and well-being of their citizens, maintaining macroeconomic stability, providing reliable infrastructure, and guaranteeing the skill and integrity of their civil servants (KKM 2006, Table B3). What, then, are aid recipients to do? Raise taxes so as to provide health care and education? Place social and political stability at risk by cutting spending? Add to the social service burden by liberalizing prices? Almost every potential solution aggravates another problem, and the KKM benchmarks thereby punish poor countries for their very poverty. If they could solve their social and economic problems, after all, they wouldn't need foreign aid in the first place. Third, and finally, KKM all but ignore the entreaties of economists who realize that "econometric work should be informed by detailed studies of individual countries" (Durlauf et al. 2005, 646). We have gone to great lengths to integrate qualitative as well as quantitative material and have therefore invoked the experiences of a

[7]KKM briefly discuss the "difficulty of estimating growth regressions" in footnote 11 (2007, 559) and ask why we abandon the random effects estimators used in our Table 1 for the fixed-effects models in our Tables 3 and 4. We note that the dependent variables in Table 1 and Tables 3–4 differ; the fixed effects in the growth models are standard tools to address the problem of unobserved country-level heterogeneity; and the—by no means uncontroversial (see Durlauf et al. 2005)—alternative estimation procedures recommended by Caselli et al. (1996) in the article cited by KKM appear to generate consistent results.

number of developing countries that have fallen short of the KKM benchmarks and nonetheless performed well in macroeconomic terms (e.g., Korea, the Dominican Republic, postwar Italy, etc.). KKM bypass the case study literature and instead compile annual data on "institutional" factors that they simultaneously portray as constant over the course of centuries. We find their oversight unfortunate. We find their data perplexing. And we can't help but think that their evidence and findings are at least in part products of their failure to fully engage the case study literature.

Political scientists have long believed that debates over concepts are at least as fertile as tests of causal inference. Laitin illustrates the point by noting that "concepts such as 'charisma' and 'the division of labor' have been longer-lasting than any valid claims about the causal effects of these concepts" (1995, 455). And conceptual innovation is not only prior to quantification but is frequently a product of qualitative research. We therefore conclude not only by underscoring the importance of conceptual rigor but by affirming the value—if by no means primacy or exclusivity—of qualitative research and, finally, by acknowledging political science's unique contributions on both fronts.

Acknowledgments

We'd like to thank Sarah Brooks, Ken Shadlen, and most importantly John Geer for valuable input.

Manuscript submitted 8 November 2006
Manuscript accepted for publication 9 November 2006

References

Ball, Terence, James Farr, and Russell Hanson. 1989. *Political Innovation and Conceptual Change*. Cambridge: Cambridge University Press.

Bardhan, Pranab. 2005. "Theory or Empirics in Development Economics?" *Economic and Political Weekly* 40 (40): 4333–35.

Caselli, Francesco, Gerardo Esquivel, and Fernando Lefort. 1996. "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics." *Journal of Economic Growth* 1 (September): 363–89.

Collier, David, and James Mahon. 1993. "Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87 (4): 845–55.

Dollar, David, and Aart Kraay. 2002. "Growth Is Good for the Poor." *Journal of Economic Growth* 7 (3): 195–225.

Durlauf, Steven, et al. 2005. "Growth Econometrics." In *Handbook of Economic Growth*, eds. Philippe Aghion and S. Durlauf. Amsterdam: North Holland, pp. 555–677.

Gerring, John. 1999. "What Makes a Concept Good? A Criterial Famework for Understanding Concept Formation in the Social Sciences." *Polity* 31 (3): 357–93.

Glaeser, Edward, et al. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271–303.

Isham, Jonathan, and Daniel Kaufmann. 1999. "The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects." *Quarterly Journal of Economics* 114 (1): 149–84.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999. *Aggregating Governance Indicators*. Typescript. World Bank.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2003. *Governance Matters III: Governance Indicators for 1996–2002*. Typescript. World Bank.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005. *Governance Matters IV: Aggregate and Individual Governance Indicators for 1996–2004*. Typescript. World Bank.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2006. *Governance Matters V: Aggregate and Individual Governance Indicators for 1996–2005*. Appendix B3. http://web.worldbank.org/WBSITE/EXTERNAL/WBI/EXTWBIGOVANTCOR/0,,contentMDK:21045419~menuPK:1976990~pagePK:64168445~piPK:64168309~theSitePK:1740530,00.html (December 3, 2006).

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2007. "Growth and Governance: A Reply." *Journal of Politics* 69 (2): 555–562.

Kurtz, Marcus, and Andrew Schrank. 2006. "Growth and Governance: Statistical Questions." Available at http://psweb.sbs.ohio-state.edu/faculty/mkurtz/papers.htm.

Laitin, David. 1995. "Disciplining Political Science." *American Political Science Review* 89 (2): 454–56.

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

Munck, Gerardo, and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices." *Comparative Political Studies* 35 (1): 5–34.

Munck, Gerardo. n.d. "Adam Przeworksi: Capitalism, Democracy, and Science." In *Passion, Craft, and Method in Comparative Politics*, ed. Gerardo Munck and Richard Snyder. Baltimore: Johns Hopkins University Press, pp. 1–63. Forthcoming.

Przeworski, Adam, Michael Alvarez, José Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990*. Cambridge: Cambridge University Press.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64 (4): 1033–53.

Wade, Robert. 2002a. "U.S. Hegemony and the World Bank: The Fight over People and Ideas." *Review of International Political Economy* 9 (2): 215–43.

Wade, Robert. 2002b. "Bridging the Digital Divide: New Route to Development or New Form of Dependency?" *Global Governance* 8 (4): 443–67.

World Bank. 2004. *World Development Report 2005: A Better Investment Climate for Everyone*. Washington, DC: World Bank.

Marcus J. Kurtz is associate professor of political science, The Ohio State University, Columbus, OH 43210. Andrew Schrank is assistant professor of sociology, University of New Mexico, Albuquerque, NM 87131-0001.

Growth and Governance: A Rejoinder

Daniel Kaufmann The World Bank
Aart Kraay The World Bank
Massimo Mastruzzi The World Bank

In this brief rejoinder we would like to point several key areas where we disagree with Marcus Kurtz and Andrew Schrank's response entitled "Growth and Governance: A Defense."

Definitions

Kurtz and Schrank (hereafter KS) begin their defense by accusing us of "abandoning" our measure of Government Effectiveness in favor of other indicators such as Rule of Law. All we noted in our original response, and reiterate here, is that we find KS's exclusive emphasis on this one particular dimension of governance to be idiosyncratic and not shared by the large economics literature on institutions and growth. We also note that we discussed all six of our governance indicators in our response simply because the critiques that KS raise of government effectiveness, notably potential respondent biases and halo effects, could equally well be applied to our other indicators. We think it important to demonstrate the scarcity of empirical support for these critiques for all six of our governance indicators.

Ideological Biases

KS remain unpersuaded by our arguments that ideological and business-oriented biases of respondents do not play a significant role in our indicators. Regrettably, however, KS continue to fail to provide any concrete evidence of such biases. KS quote the opinions of some scholars that some of our data sources are "ideologically loaded." As we noted in our response, in the case of ideological biases, we have empirically investigated whether supposedly "right-wing" think tanks give better scores of governments who share their political orientation and found virtually no evidence of this. Any empirical evidence to the contrary that KS could bring would be a welcome addition to the debate, but simply citing the assertions of others strikes us as unhelpful.

The evidence KS report from the Afrobarometer surveys on businesspeoples' preferences is irrelevant to their critique of our indicators. We acknowledge that KS have a useful suggestion that one can use the identity of respondents to this survey (who are asked to self-identify as businesspeople or other types of respondents) to investigate whether businesspeople's views differ markedly from others' on what constitutes good governance. All KS have shown is that on average businesspeople respond more favorably to questions about government services. While this is interesting, it is irrelevant to the issue at hand. What matters for our indicators is how countries are ranked relative to each other. Therefore, KS would need to show that a ranking of countries based on businesspeoples' responses differs significantly from a ranking of countries based on other responses. It is entirely possible that businesspeople on average in all surveys provide more positive responses than other respondents (perhaps they are on average just more optimistic?). But this need not affect the relative ranking of countries, which is what matters for our indicators.

KS also emphasize the relatively low correlation of the Latinobarometer survey question we use in the Government Effectiveness Indicator with corresponding variables from commercial-risk rating agencies. It is useful however to recognize that the particular Latinobarometer question we use for Government Effectiveness is an unfortunately vague question about respondents' overall trust in government. In contrast,

The Journal of Politics, Vol. 69, No. 2, May 2007, pp. 570–572

ISSN 0022-3816

the Afrobarometer survey asks several questions much more specifically focused on access to public services that are much closer to those asked of firms in the Global Competitiveness Report survey, and these two sources accordingly are much more highly correlated. Indeed, if we compare a more specific question about trust in police from Latinobarometer, that we use in our Rule of Law indicator, with a similar question about the police from the Global Competitiveness Survey, we find a quite high correlation of .77. We are therefore neither surprised nor particularly concerned at the lower correlation of the one particular—and unfortunately vague—Latinobarometer question that KS emphasize in their Defense.

We do not dispute that data sources capturing the views of businesspeople and commercial-risk rating agencies play a prominent role in our governance indicators (although certainly not an exclusive role, given our reliance on citizen surveys, NGOs, and multilateral organizations). We note also that this feature of our indicators is dictated by data availability—there simply are not many regularly updated cross-country household surveys that we can incorporate in our indicators. However, the key issue is whether country rankings based on such alternative and yet-to-be-created data sources would be substantially different from those we present. On this key point KS provide no new evidence.

Halo Effects

We can only read KS's Defense on this issue as a conceding our point that their original work failed to provide robust empirical evidence of halo effects. KS nevertheless wonder why we "depart from past practice" by using secondary school enrollment rates to expand the sample of countries (selectively citing one or two of our earlier papers on completely different topics) and suggest that the lack of robustness of their results in fact constitutes a critique of our indicators. In response, we briefly note that:

- Although of course conceptually distinct, secondary schooling rates are extremely highly correlated with stocks of human capital, with a correlation in our sample of .85. While we are unaware of any well-specified theoretical reason to prefer the stock variable over the flow variable as a determinant of government effectiveness, it is perfectly reasonable to use the flow data as a proxy for the stock in the interests of expanding country coverage. And as we showed, in this larger sample there is no systematic evidence of halo effects.
- As KS themselves note, enrollment rates could well be interpreted as an indicator of government effectiveness itself. This is a useful observation that in fact provides a further justification for using this variable, since it serves as an "objective" control for unobserved true government effectiveness. But under this interpretation, KS wonder why enrollment rates are not more strongly correlated with government effectiveness. The answer is straightforward, as the regressions KS propose also include per capita income together with schooling variables as explanators of government effectiveness. Unsurprisingly these two right-hand-side variables are very highly correlated at around .8, and this multicollinearity problem makes it difficult to identify a significant *partial* correlation between schooling and government effectiveness. In fact, the simple correlation between the two is a very respectable .72.

Growth and Governance

KS assert that our measure of Government Effectiveness should by construction be positively correlated with short-run subsequent growth, claiming that our measure contains several questions about anticipated short-run growth. This claim is based on an overly literal reading of the criteria used by just one of our 17 data sources for Government Effectiveness. DRI, a commercial risk-rating agency, uses a somewhat peculiar phrasing when it provides numbers which show their assessment of the likelihood of certain "risk events" like "a decrease in government personnel quality that lowers growth in a 12-month period." We have discussed this question at length with staff of DRI and have come to the understanding that this is simply peculiar phrasing serving their own marketing purposes and that they are actually providing an assessment of levels of civil service quality. This, combined with the sensible observation that we do not think commercial risk-rating agencies have perfect foresight about growth even one year hence, means that we are completely unconcerned that this particular measure, or the aggregate indicator, are uncorrelated with subsequent very short-run growth in the handful of empirical specifications that KS offered in their article.

We also think it important to correct the mischaracterization of the empirical literature on institutions and growth that KS offer in their defense. In particular:

- KS claim that the idea that differences in per capita income today primarily reflect differences in long-

run growth performance requires "heroic assumptions about the distribution of income across countries in 1800." This stylized fact is not an assumption, but rather is based on the careful studies by economic historian Angus Maddison, who documents as carefully as possible that the ratio of per capita incomes between the richest and poorest countries in the world 200 years ago was on the order of three or four to one, while today it is an order of magnitude greater. Based on this factual observation it logically follows that the much larger cross-country income differences we observe today primarily reflect very long-run differences in growth performance.

- KS claim that using perceptions of expropriation risk is using a measure of policy (the decision to expropriate) instead of an institution. This is a rather superficial interpretation and would only be true if studies used actual expropriations, rather than assessments of the likelihood of expropriations averaged over a decade. Investor perceptions of expropriation risk are widely interpreted as capturing something about the respect of the state for private property, which is a fundamental notion of institutional quality.
- KS dispute the use of settler mortality rates as an instrument for institutional quality because settlers were "the most violent expropriators in world history." It appears to us that KS are choosing to miss the point of one of the most influential empirical papers on institutions and growth. That Europeans propagated all sorts of atrocities in the countries where they settled is neither in dispute, nor is it in any way relevant to the empirical strategy at hand. What matters is that the disease environment faced by settlers provides an exogenous source of variation in the incentives that settlers faced to set up the institutions of property rights protection, which persist to this day. This exogenous variation can be used to statistically identify the causal effect of institutions on very long-run growth.

As we noted in our original response, we do not wish to argue that the empirical literature on institutions and growth is conclusive—and so we cannot agree with KS's characterization of us as taking an "intolerably conservative approach to social scientific argumentation." We think it is entirely appropriate for KS, or any other scholars, to critique, and so develop, the literature on institutions and growth. But we do not think that offering a few regressions, so far removed from the best-practice frontier in this literature, constitutes a serious critique that helps to advance our understanding of these important questions.

Acknowledgment

The opinions expressed here are the authors' and do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

Manuscript submitted 13 December 2006
Manuscript accepted for publication 14 December 2006

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi work at the World Bank Group, Washington, DC 20433.