Public Management and Government Performance: An International Review

Melissa Forbes
Doctoral Student
Ford School of Public Policy
University of Michigan
440 Lorch Hall
611 Tappan Street
Ann Arbor, MI 48109-1220
Tel: 734-764-3490
mkforbes@umich.edu

Carolyn J. Hill
Assistant Professor
Georgetown Public Policy Institute
Georgetown University
3520 Prospect Street NW, 4th Floor
Washington, DC 20007
Tel: (202) 687-7017
cjh34@georgetown.edu

Laurence E. Lynn, Jr.
George H. W. Bush Chair and Professor of Public Affairs
George Bush School of Government and Public Service
Texas A&M University
1064 Allen Building
4220 TAMU
College Station, TX 77843-4220
Tel: (979) 458-8033
llynn@bushschool.tamu.edu

June 26, 2005
INTRODUCTION

Performance is seemingly an obsession with governments around the world. As Frederickson and Smith (2003, 208) point out, “[a]ccountability for conducting the public’s business is increasingly about performance rather than discharging a specific policy goal within the confines of the law.”

Evidence for “this general advocacy of a performance orientation” (Pollitt and Bouckaert 2004, 126) is found, for example, in the European Commission’s commitment to “more efficient, performance-oriented working methods” (EC 2000, 8), in the British government’s widespread use of performance targets (James 2001) and in U.S. President George W. Bush’s “management agenda” and, in particular, in the Office of Management and Budget (OMB) Program Assessment Rating Tool (PART), instituted in 2002. This tool represents the most recent effort by the U.S. federal government to increase the emphasis on performance in government programs and agencies, although the 1993 Government Performance and Results Act remains in effect. It was implemented by the Bush Administration as an explicit accountability strategy:

With so much attention being paid to performance by policy makers, the public management literature’s neglect of relationships between management and performance came under mounting criticism. Peters and Savoie (1998) indicted the literature for being overly
descriptive in regard to performance and only comparing measures of performance between
countries. Additionally, Pollitt (2000) noted that very little effort has been devoted to rigorous
empirical verification of claimed results or to the identification of causal relationships underlying
them. Boyne et al. (2003, 2) asserted that “the academic community has not taken seriously the
need to evaluate public management reforms.” Fortunately, public management scholars have
begun to devote greater effort to determining how best to measure and achieve improved
performance. Of particular note is the work of Pollitt (2000), Pollitt and Bouckaert (2004),
Boyne (2003), Boyne et al (2003), Hill and Lynn (2005), and Forbes and Lynn (forthcoming) all
of which synthesize research findings from individual studies in various countries in an effort to
identify general relationships between public management and governmental performance.

This chapter synthesizes the findings of analyses by Hill and Lynn (2005) and Forbes and
Lynn (forthcoming) (hereafter referred to as HL and FL) of what public management scholars
are studying, how they are modeling causal relationships between management and performance
and, selectively and collectively, what they are finding. For the purposes of this paper, we define
government performance as the character and consequences of service provision by public
agencies. The logic of governance framework utilized in this chapter (and explained below)
considers performance measured in terms of government/public sector outputs; of markets,
firms, or private sector outcomes; and of outcomes for individuals, groups, and societies.

Of special interest is the striking tendency in American and international empirical
literatures toward hierarchical explanations of public service delivery and of the consequences of
public policies and programs, a finding at variance with the view of governance as increasingly
networked and associational rather than traditionally hierarchical. Following discussion of the
EVALUATING RESEARCH ON PUBLIC MANAGEMENT AND PERFORMANCE

A comparative, international view of public management and performance necessarily begins with a review of relevant empirical literature. A number of questions immediately arise. Should studies included in the review meet certain methodological standards? Should they be grouped by type of service such as human services or regulatory activities? Should they be restricted to certain types of functional activities such as implementation or contracting? How should “public management” and “performance” be defined? Clearly, no one right answer to each question exists. Rather, a research synthesis may differ with regard to focus, goals, organization, and other characteristic features (Cooper 1988).

In contrast to either literature reviews commonly seen in individual research papers, which appropriately focus on the specific substantive topics considered in the paper, or to more formal meta-analyses, which use statistical methods to summarize the findings of a number of individual studies (Glass 1976), our goal in HL and FL is to provide a synoptic view of the research on public sector governance from U.S. and international (primarily European, English-language) sources. We describe the types of governance relationships that researchers are investigating across a wide range of disciplines, subdisciplines, and fields. Our goal is not to describe “what works and how it works” but rather to describe “how researchers understand what works and how it works.” This fundamental step needs to be taken before considering “what works” and is especially important in the context of international comparisons.
A number of options for pursuing this goal exist. One is to classify and describe research according to the methodology employed. Public governance has been analyzed using experimental and nonexperimental methods and, within the nonexperimental category, both quantitative and qualitative methods. Within each of these categories are a number of specific approaches and analytical techniques. For example, included in the broad class of nonexperimental quantitative methods are instrumental variable or time series models. Included in qualitative methods are observational or interviewing methods. Summarizing the methods across a number of studies can provide insight into the epistemology and comparability of research findings and, therefore, into one aspect of how researchers understand what works and why. Such a synthesis does not, however, report on the substance of the underlying research.

A second option, then, is to go straight to the subject matter by classifying and describing research that originates within a particular discipline, such as economics or public health. A third and related option is to classify and describe research that pertains to a particular substantive feature such as implementation or organization type such as nonprofit organizations. These options share the drawback that they unnecessarily restrict the breadth of comparison. For example, research on education occurs in departments of economics, education, human development, political science, public administration, public management, public policy, and sociology, and appears in various disciplinary and field journals. Similarly, research findings in the context of one type of setting or organization may have parallels in other areas. While these options partly cover substantive issues, neither addresses the topic of how these issues can be understood across substantive fields.

Yet another option, which we have followed in the work reported below, is to select studies using a particular conceptual view of governance. Such a strategy would not be
appropriate for a synthesis that sought, for example, to produce a statistical estimate of an overall
effect size associated with a specific intervention. However, our strategy fits our goal of
understanding how researchers conceptualize and understand what works, and how it works,
placing these questions in the wider conceptual construct of public governance.

A Logic of Governance

A common theme in the work of governance scholars is that the rule of law — including
lawmaking, its adjudication, and its institutional expressions — is a useful starting point for
analyzing governance and interpreting relevant empirical research. Underlying this notion is
recognition that governance involves means for achieving direction, control, and coordination of
individuals or organizational units on behalf of their common interests (Vickers 1983; Wamsley
1990; Lynn, Heinrich and Hill 2001). From this starting point, it is possible to construct an
analytic framework that provides conceptual order to the systematic empirical study of
governance.

Public sector governance has been defined as “regimes of laws, rules, judicial decisions,
and administrative practices that constrain, prescribe, and enable the provision of publicly
supported goods and services” through formal and informal relationships with agents in the
public and private sectors (Lynn, Heinrich and Hill 2000a, 2000b, 2001, 7). Any governance
regime is the outcome of a dynamic process that can be summarized in terms of a core logic.¹
This process links several aspects of collective action and may be expressed in the following set
of hierarchical interactions that has been characterized by others as a “chain of delegation”
(Lupia and McCubbins 2000; Strøm 2000):

¹ For a fuller development of the ideas in this section, see Lynn, Heinrich, and Hill
between (a) citizen preferences and interests expressed politically and (b) public choice expressed in enacted legislation or executive policies;

between (b) public choice and (c) formal structures and processes of public agencies;

between (c) the structures of formal authority and (d) discretionary organization, management, and administration;

between (d) discretionary organization, management, and administration and (e) core technologies, primary work, and service transactions overseen by public agencies;

between (e) primary work and (f) consequences, outputs, or results;

between (f) consequences, outputs, or results and (g) stakeholder assessments of agency or program performance; and

between (g) stakeholder assessments and (a) public interests and preferences.

This general approach is not new to the study of public governance. For example, Kiser and Ostrom (1982) distinguished among “three worlds of action”: constitutional choice, collective choice, and operational choice. These refer, respectively, to decision-making about rules that guide choice at the collective level, decision-making about rules that guide choices at the operational level, and decisions associated with frontline work or the day-to-day activities of “street-level” administrative practice. Toonen (1998, 235) points out that “[t]he three worlds approach…opens up an understanding of public administration that goes very much beyond the organization. The three levels do not refer to different layers within a formal structure. Rather, they have to be understood as nested systems and subsystems of public policy, administrative behavior, or institutional macro structures.”

The approach is also not unique to the study of U.S. public governance. Referring to parliamentary democracies as well as presidential democracies, Strøm (2000, 266; cf. Lupia and
McCubbins 2000) describes a “chain of delegation” in “contemporary democracies, from voters all the way to civil servants that ultimately implement public policy…in which those authorized to make political decisions conditionally designate others to make such decisions in their name and place.”

The logic of governance described above, and the one that we employ in our syntheses of U.S. and international research, is consistent with, but more differentiated than, these other similar frames. In general, reference to a logic of governance encompassing interactions across different levels is useful in the design and in the synthesis of empirical research because it serves as a reminder of the endogeneity of complex governance processes and because it assists in integrating the findings of dispersed but conceptually-related literatures. Toonen (1998, 248) argued that the Kiser and Ostrom’s three-worlds “framework clarifies, for example, that reform and change at one level of analysis presupposes certain conditions at other levels of analysis.” Stated differently, reference to a wider logic or framework makes explicit the types of factors that are “presupposed” by the researcher, but that might in fact affect the relationships under study. Another advantageous feature of the three-worlds approach, in Toonen’s (1998) view, is that it can accommodate “more subtle and differentiated conceptualizations which allow us to go somewhat deeper into the actual operation of the system instead of simply scratching the surface” in comparative public administration research (237). The logic of governance just outlined provides even further leverage for assessing literature across space, time, method, and substance.

Syntheses of U.S. and International Governance Research

HL and FL employed the logic of governance to map the terrain of U.S. and international public governance research. The HL analysis focused on research concerning U.S. public
governance, drawing on over 800 articles published in over 70 academic journals covering the
twelve-year period from 1990 to 2001 (inclusive). The FL analysis focused on research
concerning international governance, drawing on over 190 articles from 51 academic journals
published between 1990 and 2004 (inclusive). All the FL articles used non-U.S. evidence; few
of the authors were American, and few of the studies were generated at U.S. universities. All
articles were published in English.

Articles were included in the FL and HL syntheses if they explicitly specified causal or
reduced form relationships between variables from two or more not-necessarily adjacent levels
of the logic of governance. Dependent variables were identified as being causally associated
with independent variables at either higher or lower levels in the logic. Studies confined to a
single level of governance were excluded from our reviews.

We used the analytic scheme summarized in Appendix A to code the dependent and
independent variables whose causal relations were under investigation in each study.\footnote{We
generally did not code the control variables in each study. In a separate but related
project, we are examining particular groups of studies in order to assess the uses and
consequences of such controls for the validity of the studies’ findings and for our ability to
synthesize findings across studies.} Thus each

study was characterized by its location within the logic of governance. The coded information
was entered into a spreadsheet that included identifiers (author, date, journal), the governance


relationships examined, the logic-of-governance codes, and the primary research method. This information constituted our databases for the FL and HL studies.

Several caveats must be noted concerning these databases. First, individual studies meet the standards of quality established by the various journals, but these standards may vary across journals and over time. Second, our strategy for selecting publications introduces three possible kinds of bias: the tendency of academic journals to publish positive findings, our own bias in favor of articles that featured a verbally or formally transparent causal model, and limitations on the journals and years of publication we included. Finally, the “correct” way to characterize a study’s variables and logic can be ambiguous: does “coordination of care” refer to efforts by treatment personnel or the strategies of their supervisors and managers? To the extent possible, we coded these according to the stated or implied definitions employed in each study.

---

3 The coding was done primarily by the authors, with occasional assistance by advanced graduate students. To ensure consistency, the coders conferred where there was any ambiguity and on subsequent discovery of any anomalies in coding.
Our reviews characterized broad patterns of research strategies and findings relating to governance, allowing us to provide a synoptic view of governance research to draw out the implications for practice and for research. Through canvassing public governance research across a wide range of disciplines and subfields and across U.S. and international research, we can discover the types of governance relationships that researchers are examining. In HL and FL, we show the full distributions of studies across the logic of governance. In this chapter, we focus primarily on the results of our syntheses that pertain to public management and performance.

PUBLIC MANAGEMENT IN A LOGIC OF GOVERNANCE

Within the multi-level logic of governance, the study of public management is concerned with managerial activity itself: the discretionary choices of actors in managerial roles, choices that are, of course, both enabled and constrained by formal authority. The need for management arises when legislation has explicitly delegated the authority to choose appropriate actions to executive agencies, when legislative mandates are ambiguous, necessitating decisions by managers as to how they should be interpreted and implemented and when fulfilling policy objectives requires managerial judgment in applying rules and standards in particular classes of cases. Because managerial discretion is virtually inevitable—few policy and service delivery domains can be completely governed by a priori rules—managerial choices are almost always a factor in government performance. But how much of a factor, under what circumstances, and compared to what?

To proceed, it will be helpful to define public management. Pollitt and Bouckaert (2004) imply that public management is concerned with adapting the structures and processes of public
sector organizations so as to ensure good organizational performance. A more elaborate version of this perspective is provided by Lynn (2003), who views public management as having three distinct but interrelated dimensions: the formal structures and processes of government, the practices and craftsmanship of individual public managers, and the taken-for-granted beliefs and values that infuse public organizations and their managers, thereby transforming them into institutions. In the HL and FL studies, public management was regarded, in the first instance, as craft activity, that is, as comprising variables whose values are determined by the deliberate choices of public managers. We then surveyed the studies in our databases to see how public management activity was conceptualized and operationalized by investigators across a range of disciplines and domains.

From the 344 studies in the HL synthesis and 88 studies in the FL synthesis that used public management as either a dependent or independent variable, three broad types of public management constructs were identified:

(1) **Administrative structures.** This category includes variables such as red tape, organizational and inter-organizational structures (such as partnership arrangements) and formalization of authority intended to constrain the behavior of subordinate and other actors. While administrative structures are not, of course, wholly determined by managerial actors—many are defined in law or by overhead organizations (finance ministries, budget offices)—the discretionary actions of public managers either create or alter structures or infuse existing structures with distinction and meaning (e.g., by creating “organizational effects” at national, state, or local office levels of administration).
Managerial tools. Within a given structural setting, managers may employ a number of different administrative instruments or mechanisms to design, implement, and evaluate policies and programs. The use of performance incentives, coordination and networking techniques, and contracts are examples of managerial tools in this category.

Management values and strategies. In contrast to structures and tools, managerial values and strategies reflect managerial choices with respect to goals, missions, priorities and adaptation to the institutional environment. Leadership, employee empowerment, inter-organizational cooperation (e.g., cooperative enforcement) and services integration, and the allocation of resources across programs and activities are all included in this category.

As shown in the “Total” row of Table 1, HL found that management variables were used almost twice as often as independent variables than as dependent variables (244 compared to 138). The ratio was slightly less, but still pronounced, for studies in the FL synthesis. Thus, both U.S. and international studies tended toward using measures of public management to explain governance phenomena, rather than considering management as a phenomenon to be explained.

[Table 1 about here]

Other interesting patterns are evident in the table. First, the distribution among the three subtypes, across use as an independent and dependent variable, is quite similar for the U.S. studies included in the HL synthesis: administrative structures are used most often, then values and strategies, and finally tools. In U.S. studies, administrative structures are even more heavily used as independent variables compared to their use as dependent variables (52 percent
compared to 43 percent). In FL’s international studies, we see a consistent—but different—pattern across independent and dependent management variables: values and strategies are used most often, then administrative structures, then tools. Moreover, values and strategies receive relatively more emphasis as dependent variables than as independent variables (63 percent compared to 47 percent).

To summarize, in terms of sheer numbers (without regard to study quality) we are amassing more information about how management affects other outcomes (usually further down the hierarchy, as will be discussed later in the chapter) than about how other levels of governance affect the administrative structures, tools, and values and strategies of managers. Moreover, U.S. and international research diverge in their patterns of emphasis on administrative structures compared to values and strategies. But the two bodies of research share the characteristic that relatively little public management research attempts to understand what are the influences of, or what influences, managerial use of tools or instruments (quality control, planning), despite their importance in prescriptive literature.

**PERFORMANCE IN A LOGIC OF GOVERNANCE**

As noted earlier, government performance can be defined broadly as the character and consequences of service provision by public agencies. The definition of performance we employ is similar to, but distinct from, that used by other studies. For example, performance for Boyne (2003) means efficiency, responsiveness, or equity. Further subcategories, which align quite closely with particular e- and f-level categorizations using a logic of governance framework, include quantity of outputs, quality of outputs, efficiency, equity, outcomes, value for money, and consumer satisfaction. Pollitt (2000) sets forth broader criteria for performance: savings,
improved processes, improved efficiency, greater effectiveness, and an increase in the overall capacity/flexibility/resilience of the administrative system as a whole.

Examined within a logic of governance, we define measures of performance at the service delivery, consequences/outcomes, and stakeholder assessment levels (see Appendix A). At the level of primary work or service transactions (e-level), performance measures are primarily conceived as changes in accountability, efficiency, costs, or quality. Performance changes at the f-level largely measure broader changes in the final outputs/outcomes of a particular program or changes in law.\footnote{The distinction between e- and f- levels can be somewhat subjective depending on how authors frame their research questions. Some variables coded as primary work measures of performance may be viewed as public sector outputs as well. However, in both analyses, the explanatory models are top-down in their orientation. Thus, the general findings of the study would not be affected by changing e-level to f-level classifications.} Finally, stakeholder assessments of agency or program performance (g-level) may themselves be considered measures of performance.

The distribution of performance variables examined in the HL and FL syntheses was quite similar: service-delivery and consequences (e-level and f-level) variables were by far the most common dependent variables overall, each used in about one-quarter to one-third of all studies examined. In contrast, these types of variables were infrequently used as independent
variables. As shown in Appendix A, f-level performance variables were classified as either outputs or outcomes, and within each of these categories, further subdivided into whether they measured results with respect to government/public sector, the private (market) sector, or individuals/society. The distribution of studies across these six subcategories was remarkably similar in the HL and FL syntheses, where the modal categories were individual/society outputs and outcomes. In contrast to the heavy use of e- and f-level dependent measures, stakeholder assessments of performance (g-level) were the least-frequently used dependent (and independent) variables in the HL synthesis, and among the least-frequently used in the FL synthesis.

Overall, the body of relevant research on performance comprises primarily longitudinal studies, usually consisting of before-and-after or interrupted time series studies that attempt to associate results or improvements with the reforms that produced them, both with and without controlling for other factors that might have mediated reform results. At the level of primary work and service delivery (e-level), FL’s review included studies that examined changes in the performance of entities like primary health centers in India (Varatharajan et al 2004); technical efficiency of power generation in China (Lam and Shiu 2001); cost savings and effectiveness of a NPM merger in Welsh mental health care (Kitchener and Gask 2003); service efficiency, 

While many variables at the e-level are considered “performance” variables, not all are. Such “non-performance” variables at the e-level (subtypes e1 through e5 listed in Appendix A) are frequently employed as independent variables to model performance changes at the f-level. Twenty-nine percent of HL studies and seventeen percent of FL studies use an e-level independent variable and an f-level dependent variable. For example, a study might test the effect of field worker/office discretion (e2) on the number of clients served (f11).
performance, and quality (Barnett and Newberry 2002); and changes in transportation service
delivery efficiency (Pina and Torres 2001). HL’s review included studies that examined e-level
performance such as school district efficiency (Grosskopf et al. 2001); initiation, completion,
and pace of Superfund site cleanup (Hird 1990); mental health service coverage (Grusky and
Adams 1994); accessibility to disability benefits (Rosenheck, Frisman, and Kasprow 1999); and
expenditures on capital, water, sewers, and highways (Nunn 1996).

Among studies that tested for performance changes at the output/outcome level (f-level),
the primary focus of research included in both the HL and FL studies was on the
individual/societal level rather than on outputs and outcomes for the public or private sectors.
Examples of performance outputs at the individual/societal level in the studies examined by HL
and FL include: homicide rates, school test performance, school dropout rates, welfare payments
and employment rates for welfare-to-work clients, households’ expenditures on food, recidivism,
use of bicycle helmets, and citizen participation in local recycling programs. Individual/societal
outcomes included in the studies included changes in workforce quality and productivity,
development of active citizenship among young people, heightened public knowledge about
health issues, earnings impacts for clients in local welfare-to-work programs, exits from poverty,
and life saving effects of seat belts.

Among studies examining g-level dependent variables, most are concerned with
stakeholder assessments of changes to government services and programs. Examples include the
privatization of services (Poister and Henry 1994, Becker, Dluhy and Topinka 2001), citizen
opinions on drugs chosen by the legislature for higher reimbursement in Finland (Vuorenkoski et
al, 2003), and the effects of client influence in program decisions on their assessment of
government programs in the U.K. (Bache 2001).
WHAT AFFECTS GOVERNMENT PERFORMANCE?

The distributions of independent variables in studies that sought to explain dependent variables at the e,f, or g-levels are shown in Table 2. We first examine the patterns for primary work variables (top panel). First, both HL and FL syntheses found very few studies that sought to explain primary work as a function of output/outcomes at the f-level, or of stakeholder assessments at the g-level: the explanations for primary work came from “above” in the logic. Second, an emphasis on structural (c-level) and management (d-level) variables was evident as explanatory factors for primary work across both U.S. and international studies. About half of the U.S. and international studies used structural variables to explain primary work dependent variables.

A sharp divergence is evident, however, across the HL and FL results with respect to the role that management-level variables play in explaining primary work: 43 percent of U.S. studies used a d-level variable to explain primary work, while 63 percent of international studies explored such a relationship. Studies in the U.S. synthesis showed a greater tendency to explore multi-level relationships within the primary work level (almost 25 percent, compared to only eight percent of international studies). With respect to explaining outputs or outcomes at the service delivery level, researchers of U.S. governance tend to rely primarily on structure explanations, management, and service delivery activities in that order. International studies tend to investigate these same types of dependent variables primarily with management-level explanations and only secondarily with structural explanations.

[Table 2 about here]
The middle panel of Table 2 reports the distribution of independent variables for f-level dependent variables. In contrast to explanatory factors for primary work just discussed, where divergence in approach between U.S. and international governance research surfaced at the managerial level, here we see a marked difference in explanatory factors occurring at the structures level. Scanning across the HL row, f-level dependent variables were modeled using either c-, d-, or e-level variables in about one-third of the studies (37, 31, and 29 percent, respectively). In contrast, the FL synthesis shows that structures are relied upon relatively more as explanatory factors, compared to management and primary work explanations (56, 18, and 17 percent respectively). While not evident from the table, separate tabulations showed that over 60 percent of the FL and HL studies employing d-level independent variables and f-level dependent variables focus on individual and societal level (rather than public or private sector) outputs or outcomes.  

As shown in the bottom panel of Table 2, significantly fewer studies in both the HL and FL databases use g-level stakeholder assessment dependent variables compared to e- and f-level studies. Due to the very small number of studies in this category, it is difficult to draw firm conclusions about modeling tendencies or patterns.

Looking across e- and f-level measures of performance, both HL and FL syntheses show that studies examining service delivery (e-level) or outputs/outcomes (f-level) often used explanatory variables at the structural level (c-level). These types of studies thus skip one or two levels of governance — management and service delivery — to explain performance changes. For example, Boex (2003) examined how central government budget transfers to local

---

6 An overwhelming majority (almost 85 percent) of these studies employ quantitative methods to study the relationship between the (d) and (f) levels.
government in Tanzania (c-level) affected the redistributive effect of local government allocations (f-level). Such patterns of inquiry assume, in effect, that intervening managerial activity or implementation did not mediate how structures affect performance.

The tendency for researchers to use formal structures of authority as explanatory variables for changes in performance raises the question of whether public management researchers are neglecting potentially significant causal processes at intermediate levels of governance. To illustrate this potential problem, one might question whether the way in which budget transfers were handled administratively in the Boex (2003) study might also affect the redistributive effect of local government allocations. In another example, one of the studies included in the FL synthesis, Smith and Hardman (2000), treated the UK’s National Literacy Strategy Program’s framework for teaching (c-level) as an explanatory variable for both teacher opinion about the program (e-level) and school exam results (f-level). An obvious question, however, is whether teachers’ attitudes toward the program that resulted from the teaching framework imposed on them could also mediate the exam performance of their students.

In any social science inquiry, some explanatory variables will undeniably mediate the effects of other explanatory variables on performance, a point Boyne (2003, 389) emphasizes as well. The job of public management researchers is to tease out these often subtle, but potentially telling, complex causal relationships between public management variables and public sector performance. The logic of governance provides public management scholars with a framework to recognize (and either accept or reject) the role of mediating variables in their analyses.

Examining the types of relationships that researchers have explored between d-level independent variables and e- or f-level dependent variables may illustrate possible missing pieces of the picture. Among international studies using d-level independent variables to explain
variables at the e- and f-levels, researchers seem particularly focused on the role management values and strategies play in determining performance (Table 3). Twenty-nine FL studies investigate the relationship between management values and strategies and e- or f-level variables compared to sixteen studies each for administrative structures and managerial tools. A typical international study using managerial values or strategies to model performance changes is Valdivia’s (2002) study of how management decisions regarding health infrastructure location has affected the utilization of outpatient health care services in Peru. In contrast, the number of d-level American studies employing managerial values and strategies and administrative structure explanatory variables to model e- and f-level changes were almost the same (83 and 88 studies respectively). Only half as many HL studies used managerial tools to model e- or f-level variables, such as Ehrenberg et al’s (1991) study of how teacher leave policies were related to student academic performance.

[Table 3 about here]

The reality of complex causal processes and mediating variables raises a difficult methodological dilemma for public management scholars. Qualitative studies (for example, Bache 2001) may have access to more detailed, in-depth data that enable systematic consideration of the mediating effects of explanatory variables and how these effects might influence performance. (Interestingly, the FL synthesis of international studies included a larger number of qualitative articles in its database than the HL synthesis.) Testing for multi-causal processes affecting performance, however, poses difficulties for both quantitative and qualitative researchers. For the quantitative researcher, sufficiently large sample sizes may enable the use of an array of statistical controls, but data limitations may constrain the types of variables and levels of analysis that can be studied. For the qualitative researcher, rich detail may be available about
many core and mediating factors, but sample size often precludes the ability to convincingly rule out confounding factors. Ultimately, the study of a phenomenon where endogeneity is as complex as public service performance will benefit from both quantitative and qualitative epistemologies.

Regardless of the methodology chosen by public management scholars, the logic of governance is a parsimonious and useful analytical tool for researchers to use when thinking about causal relationships related to performance. At the beginning of the research design process, it provides a systematic checklist for thinking about how explanatory variables operate in reality to affect service provision and outcomes. As the researcher proceeds, the logic of governance can also operate as a check on modeling causal relationships that neglect the mediating effect of variables that fall between explanatory variables and the dependent variable in the logic of governance. Finally, the logic of governance assists in the interpretation of findings, enabling investigators to speculate more precisely on how omitted variables might have influenced their findings, making explicit what is “presupposed.”

CONCLUSION

As noted at the beginning of this chapter, a consensus exists among researchers that relatively little is known about determinants of performance. With the almost 1,000 articles in our databases of U.S. and international governance research, we are still not able to say for certain what works, for whom and under what conditions. Yet as Cooper and Hedges (1994, 4) point out, when research results differ, one “should not pretend that there is no problem or decide that just one study, perhaps the most recent one, produces the correct finding. If results that are expected to be similar show variability, the scientific instinct should be to account for the
variability by further systematic work.” By describing how researchers of U.S. and international public governance are trying to understand governance and performance, we can assess whether we as a research community are asking the right kinds of questions, where more detailed syntheses might be informative, and where further research and replication of results is needed.

While the United States is not exactly an outlier among nations, public governance in the United States is widely regarded as unique in the extent and influence on policy making and public management of its formal separation of powers and of its individualistic orientation. Yet the findings of the FL and HL analyses regarding how researchers understand what works and why both imply that the determinants of government performance are multifarious and are to be found at multiple hierarchical levels of governance that are interrelated in complex ways.

We also noted some key differences between approaches to the study of U.S. and international governance research. International investigators exhibit somewhat different modeling strategies, tending, for example, to favor more linear managerialist hypotheses—changes in structure lead to changes outcomes, for example—than American research, which is more concerned with inter-level complexities and the polycentric nature of governance. A possible justification for this difference is that American public management, heavily influenced by organized interests, the diffusion of power, and legislative and judicial micro-management, is in fact more constrained and polycentric than is the case in more unitary or statist regimes and should be expected to exhibit more complex patterns of outcome determination.

A particularly interesting substantive finding of both the FL and HL studies, as seen in Table 2 but exhibited for other levels of dependent variables as well, is the clear tendency toward hierarchical explanations of primary work and the consequences of government action. In the

---

7 See, for example, Lijphart (1999), Hofstede (2001), and Pollitt and Bouckaert (2004).
literature, hierarchy is the backbone of governance. This is not to say that hierarchical influences are necessarily decisive or even particularly effective. Evidence on the effectiveness of the chain of command is, at best, mixed. Nonetheless, our finding is notable because it is at considerable variance with the view of governance popular in both European and American literature as increasingly networked and associational. HL speculate that the widely-touted “paradigmatic” shift away from hierarchical government toward horizontal governing (hence the increasing preference for “governance” as an organizing concept) is less pervasive than is supposed and that it is usually tactical: polyarchic tools and administrative technologies are being employed, perhaps increasingly so, to facilitate public governance within constitutionally and financially hierarchical regimes. The work of Hall and O’Toole (2000, 2004) supports this view (at least at the federal level in the U.S.). Their analysis of U.S. laws and regulations spanning approximately 30 years found numerous instances of multi-actor relationships, but no marked increase in networked relationships (at least as codified in laws and regulations) over this period.

As Frederickson and Smith (2003, 224) note, “hierarchy is necessary for conjunction to exist” because the American political system remains hierarchical and jurisdictional. The “chain of delegation” also characterizes parliamentary democracies (Strøm 2000). And, when it comes to answering multi-level “why” questions, the evidence suggests that hierarchy preoccupies field researchers. The fact that relatively few studies in either database examined more complex patterns of causality incorporating the configurational, endogenous nature of governance may reflect the paucity of data, which constrains modeling efforts to postulating more straightforward, linear causality. It may well suggest something more revealing, however: conjectures by hundreds of investigators in specialized domains that the interesting questions of administration and management concern the effects of hierarchical interactions more than of
horizontality. We cannot rule out that researchers are investigating these questions (instead of other questions) due to data constraints, but the consistency of research agendas across policy domains and intellectual subfields is suggestive of the kinds of questions, and answers, that are of interest to the audiences for empirical research throughout the worldwide public management community.

This issue warrants much further investigation, however. Achieving greater insights into the interactions between hierarchical authority and the interdependence of the many public- and private-sector agents engaged in service delivery, as Provan and Milward do in this volume, would greatly enlighten our understanding of delegation and accountability, the efficacy of relying on polycentric arrangements to accomplish policy mandates, and the results of so-called “post-bureaucratic” governance arrangements on public service performance and the satisfaction policy makers and citizens derive from it. While the data and conceptual demands of such research are often daunting, the complementarities and tensions between hierarchy and networks are a cutting edge issue for the public management field.
References


### Table 1
Public Management in HL and FL Syntheses

<table>
<thead>
<tr>
<th>Subtype of Management Variable</th>
<th>Studies that Use Public Management as a(n) Independent Variable</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hill &amp; Lynn</td>
<td>Forbes &amp; Lynn</td>
</tr>
<tr>
<td>Administrative Structures</td>
<td>127 (52%)</td>
<td>22 (29%)</td>
</tr>
<tr>
<td>Managerial Tools</td>
<td>52 (22%)</td>
<td>18 (24%)</td>
</tr>
<tr>
<td>Management Values &amp; Strategies</td>
<td>104 (43%)</td>
<td>36 (47%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>244</strong></td>
<td><strong>76</strong></td>
</tr>
</tbody>
</table>

**Sources:** Hill and Lynn (2005), Table 4; Forbes and Lynn (forthcoming), Table 5.

**Notes:** Cells show the number of studies, then the column percentages for the studies in the HL and FL syntheses. For example, of the 76 studies in FL that examined a management-level independent variable, 36 of them, or 47 percent, used a type of public management variable that could be classified as “management values and strategies.” Column percentages do not sum to 100 percent, because some studies used more than subtype of management variable.
Table 2
Explanatory Factors of Performance in HL and FL Syntheses

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level of Independent Variable Used to Model the Dependent Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a) (b)</td>
<td>(c) (d)</td>
<td>(e) (f)</td>
<td>(g)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e) Primary Work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill &amp; Lynn</td>
<td>40 (17%)</td>
<td>35 (15%)</td>
<td>125 (54%)</td>
<td>100 (43%)</td>
<td>56 (24%)</td>
<td>7 (3%)</td>
<td>2 (1%)</td>
<td>232</td>
</tr>
<tr>
<td>Forbes &amp; Lynn</td>
<td>1 (2%)</td>
<td>7 (14%)</td>
<td>25 (49%)</td>
<td>32 (63%)</td>
<td>4 (8%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>51</td>
</tr>
<tr>
<td>(f) Outputs/Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill &amp; Lynn</td>
<td>20 (7%)</td>
<td>76 (26%)</td>
<td>111 (37%)</td>
<td>91 (31%)</td>
<td>85 (29%)</td>
<td>7 (2%)</td>
<td>1 (&lt;1%)</td>
<td>298</td>
</tr>
<tr>
<td>Forbes &amp; Lynn</td>
<td>5 (6%)</td>
<td>15 (18%)</td>
<td>47 (56%)</td>
<td>15 (18%)</td>
<td>14 (17%)</td>
<td>1 (1%)</td>
<td>2 (2%)</td>
<td>84</td>
</tr>
<tr>
<td>(g) Stakeholder Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill &amp; Lynn</td>
<td>9 (39%)</td>
<td>0 (0%)</td>
<td>7 (30%)</td>
<td>3 (13%)</td>
<td>4 (17%)</td>
<td>7 (30%)</td>
<td>2 (9%)</td>
<td>23</td>
</tr>
<tr>
<td>Forbes &amp; Lynn</td>
<td>1 (7%)</td>
<td>3 (21%)</td>
<td>6 (43%)</td>
<td>6 (43%)</td>
<td>4 (29%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>14</td>
</tr>
</tbody>
</table>

Sources: Hill & Lynn (2005), Table 1; Forbes and Lynn (forthcoming), Table 1.
Notes: Cells show the number of studies, then the row percentages for the studies in the HL and FL syntheses. For example, of the 232 studies in HL that examined an e-level dependent variable, 100 of them, or 43 percent, modeled that dependent variable using a d-level independent variable. Row percentages do not sum to 100 percent, because some studies used more than one level of independent variable to model the dependent variable.
Table 3
How Does Public Management Influence Performance?

<table>
<thead>
<tr>
<th>Level of Dependent Variable</th>
<th>Type of d-level Independent Variable</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Administrative Structures</td>
<td>Managerial Tools</td>
<td>Management Values &amp; Strategies</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Hill &amp; Lynn</th>
<th>Forbes &amp; Lynn</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(e) Primary Work</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill &amp; Lynn</td>
<td>45</td>
<td>24</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forbes &amp; Lynn</td>
<td>10</td>
<td>13</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(f) Outputs/Outcomes</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill &amp; Lynn</td>
<td>43</td>
<td>20</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forbes &amp; Lynn</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subtotal Hill & Lynn  88  44  83  
Subtotal Forbes & Lynn 16  16  29  
Total 104  60  112

Sources: : Hill & Lynn (2005), Table 6; Forbes & Lynn (forthcoming), Table 7.  
Notes: Cells show the number of studies of that type in the HL and FL syntheses.  For example, 45 studies in the HL synthesis used a d-level explanatory measure of administrative structures to explain an e-level variable.
Appendix A
A Logic of Governance Identifies Relationships…

between  (a) citizen preferences and interests expressed politically, further disaggregated into:
(a1): primordial citizen preferences and interests;
(a2): private firms, organizations, behavior, participation, etc.; and
(a3): interest groups

and

(b) public sector decisions, activity, and influence, which may be further disaggregated into
(b1): legislator preferences expressed in action or in enacted legislation;
(b2): executive policies and, in a federal system, federal-level policies and influence; and
(b3): court decisions;

between  (b) public sector influence, activity, and choice

and

(c) formally authorized structures and processes of public agencies at federal or state level, including regulatory authority, disaggregated into

(c1): hierarchy/structure
  (c11) type of ownership
  (c12) level/type of government
  (c13) internal government entities
  (c14) political atmosphere

(c2): mandated behavior
(c3): policy design and elements
(c4): fiscal situation
(c5): other

between  (c) the structure of formal authority

and

(d) the de facto or discretionary organization and management of the executive branch or of executive agencies, programs, and administrative activities, disaggregated into

(d1): administrative structures;
(d2): tools;
(d3): values and strategies;

(continued on next page)
between (d) discretionary organization, management, and administration

and

(e) primary work or service delivery activities of public agencies (the availability, type, quality, and cost of publicly sponsored goods and services); which may be disaggregated into:

(e1): program design features
(e2): field worker/office beliefs and values
(e3): administrative processes and policies
(e4): work/treatment/intervention
(e5): client influence, behavior, and/or preference
(e6): use of resources and/or performance (i.e., efficiency, costs, quality, etc.)

between (e) primary work activities/transactions

and

(f) consequences, outputs, outcomes, or results, which may be further disaggregated into:

(f1): outputs
   (f11): government/public sector
   (f12): market/firm/private sector
   (f13): individual/society

(f2): outcomes
   (f21): government/public sector
   (f22): market/firm/private sector
   (f23): individual/society

between (f) consequences, outputs, outcomes, or results

and

(g) stakeholder assessments of agency or program performance (i.e., judgments about whether government is “working” that motivate them to political action);

between (g) performance assessments expressed politically

and

(a) public and private interests and preferences