

when deciding how to allocate time and money. The first involves the maintenance and adaptation of existing structures and routines (M1). The second entails the development and implementation of a conscious strategy for managing the environment in which an organization operates (M2), which in turn takes the form of a balance between seeking to exploit environmental changes and to defend the organization from those changes (M3/M4) (O'Toole and Meier 1999).

Table 1 lists the main management approaches associated with the three theoretical perspectives on local government management and performance that we have introduced, and which we identified as having been the focus of one or more of the empirical studies reviewed. Table 1 also indicates how each approach to local government management can be incorporated within Meier and O'Toole's model of public management, and the anticipated relationship between those different approaches and local government performance. This conceptual framework forms the basis for our analysis of the studies of local government under review. The management approaches included in the final analysis are indicated in the right-hand column of the table.

METHODS

The empirical literature on management and organizational performance in local governments was located in the Anglophone public administration journals listed in the Thompson Reuters Web of Science Social Sciences Citation Index (SSCI). Not all of the abstracts reviewed indicated that a study was situated in local government. Rather than using search terms, we reviewed each article by hand from 1970 to 2012. This search

Table 1
Local Government Management Approaches

Theory	Key Concepts	O'Toole and Meier	Anticipated Relationship with Performance	Included in the Analysis
Economic theory	Organization size	M1, M3/M4	+	✓
	Contracting out	M3/M4	- or +	
	Competition	M3/M4	- or +	
	Collaboration	M3/M4	+	
	Coproduction	M3/M4	+	
Contingency theory	Administrative intensity	M1	-	
	Centralization	M1	-	
	Integration	M1	+	
	Strategy content	M3/M4	+	✓
	Planning	M1	+	✓
Resource-based theories	Management systems	M1	+	
	Staff quality	M1	+	✓
	Personnel stability	M1	+	✓
	Leadership	M1	+	
	Human resource management practices	M1	+	
	Representative bureaucracy	M3/M4	+	✓
	Networking	M3/M4	+	✓

procedure resulted in 490 empirical articles featuring questions of management and organization. We examined these studies in greater depth, and implemented the following decision rules for inclusion in the review: the unit of analysis was an organization, or part thereof (thus excluding studies at the individual level of the manager or citizen); studies included measures of management and organization (thus excluding studies focused solely on management or performance); performance was operationalized as the dependent variable; and articles presented statistical results in the form of multiple regressions that could be used in the “support score” meta-analytic technique used in this study (see below).² This procedure resulted in a sample of 86 articles, containing 999 independent tests of some aspect of the management-performance relationship.³

The review strategy that we adopted benefits from focusing on peer-reviewed journal articles that were judged to be of suitable quality for publication by editors following a blind review process, and therefore expected to meet the basic requirements of theoretical and methodological rigor. It does, however, exclude unpublished articles on management and performance in local government and work sponsored by government, national and global organizations, such as the OECD, with an interest in the achievements of public organizations, along with books and book chapters. This approach may lead to bias, by overstating the relationship between management and performance if articles that contain statistically significant results are more likely to be published. Even so, estimates from other fields suggest that the magnitude of the bias is likely to be small (Rosenthal 1991).

The first group of studies we identified were published in the 1970s (Davies, Barton, and McMillan 1971; Hansen and Kjellberg 1976). Three more were published in the 1980s, and eight more in the 1990s. Forty-eight were published in the 2000s and 25 in 2010–2012. The distribution of these articles over time suggests that interest in this topic soared in the 2000s, reflecting the upsurge in interest in questions of performance during that period. The majority of the studies were conducted in the United States (49), with 28 based in the United Kingdom (England and Wales), 6 in other European countries, and one each in Israel, Pakistan, and South Korea. Local governments in most countries vary in scope and purpose. The studies in this sample capture these variations with 54 studies examining single-purpose governments including fire brigades, police departments and school districts, and the balance (32) investigating multipurpose governments that deliver several types of public service. Because most of the studies were conducted in the United States and the United Kingdom, evidence of the provision of acute and primary healthcare by local governments is largely absent from our sample.

All of the studies were focused on a single set of organizations with an average sample size or number of observations of 1,070, ranging from 40 (Andrews et al. 2011) to 6,994 (Pitts 2007). Seventy percent of the studies (70.5%) implemented research designs that sought to address endogeneity and tease out causality, largely by introducing some semblance of time into the regression model: 42 used a panel design and 19

2 Independent or linear associations between management and performance are considered in this review. Nonlinear, mediated, and moderated relationships are examined in some empirical studies (e.g., Andrews and Boyne 2011; Meier and Bothe 2000; and Walker et al. 2011), but not in sufficient numbers to undertake a comprehensive review on this occasion.

3 The full list of citations are included in the references, and identified by use of an asterisk.

others a lagged data structure, leaving 31 purely cross-sectional studies.⁴ Within these studies, eight dimensions of performance were used: effectiveness measures were most frequently used (459), followed by equity (148), aggregated performance indices (128), efficiency (79), service quality (68), customer or user satisfaction (64), cost effectiveness (31), and output quantity (23). Measures of management were evenly balanced between archival (497) and perceptual (502) methods of operationalization, whereas measures of performance tended toward the use of archival (or administrative) data sources (793), perceptual (or survey based) (163), and indices incorporating both (43).⁵

Support Score Method

The method used to combine and synthesize the results of the empirical evidence is based on the percentage of statistical tests that support the hypothesis that management positively or negatively (see [table 1](#)) influences performance ([Light and Smith 1971](#)). We use the support score approach because the majority of the studies reviewed implemented multiple regression techniques without reporting correlations ([Boyne 2003](#); [Damanpour 2010](#)).⁶ To count as support for the hypothesis, two conditions must be satisfied. First, the results must be in the predicted direction. Second, the results must be statistically significant; that is, greater than would be likely to arise by chance.⁷ By applying these criteria to all of the tests in a single study, a support score can be calculated as a percentage of all of the tests reported in the study.

Following this, an aggregate support score can be calculated across all of the studies in at least two ways ([Boyne 2002b](#); [Rosenthal 1991](#)). First, the support score for each study can be treated equally, regardless of whether it contains 1 or 300 tests. Second, each study can be weighted (multiplied) by the number of tests in that study, with equal weight attached to each test rather than to each study. The advantage of the weighted mean is that studies that only report a small number of tests do not have a disproportionate influence on the analysis, whereas the advantage of the unweighted mean is that studies that conduct a large number of tests on the same data set are not given undue importance. An examination of the number of tests in each study shows that there is some right-side skew, with a number of studies reporting in excess of 40 tests over the average of 12 (standard deviation [SD] 13, minimum 1 and maximum 63). These studies are essentially outliers and could affect the robustness of the results, typically by suppressing the support score and producing heightened discrepancies

4 [Hill \(2005\)](#) included a panel and a lagged dataset, [Walker and Boyne \(2006\)](#) included cross-sectional and lagged analysis.

5 A document summarizing the different indicators used to measure each of these dimensions of performance can be obtained from the authors.

6 Alternative statistical and computational approaches can be adopted for the purposes of meta-analytic reviews. These techniques generally require the reporting of correlation matrices. Unlike studies using correlation coefficients, the support score method reports statistically significant associations from regression models that control for other variables, thereby reducing concerns about bias arising from spurious relationships ([Damanpour 2010](#)).

7 In the social sciences, authors sometimes report statistically significant results at $p < .1$, a practice adopted by some authors in this review. We include test results at 0.1 in the analysis. Sensitivity tests indicated that the use of $p < .1$ as against $p < .05$ had little effect on the support scores.

between the unweighted and weighted scores. Given these unusual characteristics in the data, we report the unweighted support score first, and conduct sensitivity tests for studies with large numbers of tests and report these in footnotes. Finally, in interpreting the support scores, we follow [Boyne \(2002b\)](#) and [Damanpour \(2010\)](#), who proposed that unweighted and weighted support scores of 50% or more shows strong support for a hypothesis, that support is moderate if one support score is above 50% and that the hypothesis is not supported when both scores fall below 50%.

MANAGEMENT AND PERFORMANCE EVIDENCE

To ensure that we focus our attention on approaches to local government management, which have been subject to sustained investigation, the support score results are presented in tabular form exclusively for those aspects of management examined in 10 or more studies.⁸ Hence, seven local government management approaches form the basis for our review of the evidence on performance effects: organization size, strategy content, planning, staff quality, personnel stability, representative bureaucracy, and managerial networking. These studies account for 80.2% of the total number of articles reviewed (69 out of 86) and 70.8% of the tests (708). Thus, the majority of the research conducted on the management and performance of local governments has been focused on these questions, rather than on the other concepts of management that we identify in [table 1](#). Nonetheless, we are keen to stress that there is work going on in these areas, although it is relatively sparse when compared with that analyzed (see footnote #8). For now, systematic evidence suitable for meta-analytic review can only be garnered from findings on the seven management approaches benefiting from a larger number of studies. Included in the following seven tables are information on the sample, country of study, purpose of the government being investigated, the operationalization of the measure of management and performance, together with the support score for each study and a total for each area of management.

Economic theories of local government production have underpinned empirical research in the field for more than thirty years, and the role of size as a determinant of organizational performance was examined in more studies than any other management approach ([table 2](#)).⁹ These studies were predominately based in the United States, examined multipurpose governments and were more likely to operationalize management and performance using archival data. A range of performance measures were used to gauge the effects of size, the most frequent being efficiency; no doubt

⁸ Support score results for those facets of local government management that were the focus of between five and nine empirical studies are available on request: administrative intensity (5 studies, 29 tests), collaboration/partnership (5, 19), competition (9, 40) and leadership (8—of which 3 examine executive succession, 28).

⁹ The concept of organizational size sits at the intersection between managerial discretion and environmental constraint. At first glance, the size of a public organization may seem to be a variable that is not within the purview of public managers; determined rather by the mission of the organization and the clientele to whom it must provide a service. Yet, the senior management in local governments may possess a degree of discretion over the size of the organization that they manage. Whether by seeking amalgamation with neighboring governments or by implementing expansionary policies to compete for fiscal migrants, discretion over the size of local governments is an important component in the manager's toolkit, however crude, for managing the environment.

Table 2
Organization Size

Study	Sample	Country	Purpose	Management		Performance		No. of tests			
				Size	Measure	Dimension	Measure	+	-	ns	
Davis, Barton and McMillan (1971)	86	United States	SP	Population	Archival	Efy	Archival	2	100	0	0
Hansen and Kjellberg (1976)	460	Norway	MP	Population	Archival	Efy	Archival	4	50	0	50
Fitzgerald and Durant (1980)	438	United States	MP	Population	Archival	US	Perceptual	1	0	100	0
Walker and Williams (1986)	70	United Kingdom	MP	Population	Archival	Q	Archival	2	100	0	0
DeSantis and Renner (1994)	1,257	United States	MP	Population	Archival	Efy	Archival	4	0	0	100
Miranda and Lerner (1995)	539	United States	MP	Population	Archival	Efy	Archival	1	100	0	0
Nunn (2001)	185	United States	SP	Population	Perceptual	Efy	Archival	4	0	0	100
Smith (2003)	179	United States	SP	Population	Archival	Q	Archival	1	0	100	0
Nicholson-Crotty and O'Toole (2004)	544	United States	SP	Population	Archival	Eft	Archival	3	0	0	100
Andrews et al. (2005)	148	United Kingdom	MP	Population	Archival	In	Archival	1	100	0	0
Melkers and Willoughby (2005)	194	United States	MP	Population	Archival	CE, In	Perceptual	2	0	0	100
Andrews, Boyne, and Walker (2006a)	120	United Kingdom	MP	Population	Archival	In	Archival	1	0	0	100
Carmeli (2006)	263	Israel	MP	Population	Archival	Efy	Archival	8	0	0	100
Goerdel (2006)	2,120	United States	SP	Student attendance	Perceptual	Eft	Archival	5	0	0	100
May and Winter (2007)	159	EU	SP	Population	Archival	Eft	Perceptual	5	20	0	80
Sorensen (2007)	175	EU	SP	Population	Archival	Efy	Archival	3	100	0	0
Andrews and Boyne (2010)	88	United States	MP	Population	Archival	In	Archival	2	50	0	50
Andrews and Boyne (2011)	148	United Kingdom	MP	Population	Archival	CE, Eq, In	Archival	3	0	0	100
Aslam and Yilmaz (2011)	1,885	Pakistan	MP	Population	Archival	Q	Archival	4	50	0	50
Andrews et al. (2012)	178	United Kingdom	MP	Population	Archival	In	Archival	2	0	100	0
						Number of studies		20			
						Number of tests		58			
						Unweighted		34	10	57	
						Weighted		26	3	71	

Note: Purpose: MP = multipurpose, SP = single purpose; Performance dimension: CE = Cost effectiveness, Eft = Effectiveness, Efy = Efficiency, Eq = Equity, In = Index Q = Quality, US = User satisfaction.

management, which argue that a rational approach to the management of organizations is associated with higher performance (Walker et al. 2010).

According to resource-based theories, attracting and retaining high-quality staff is critical to organizational success (Barney 1991). The stronger the talent within an organization, the more likely that it can be relied upon or marshaled to achieve higher levels of performance. The evidence for this thesis is strong. Table 5 indicates that around three-quarters of the studies examining staff quality and local government performance support the hypothesis that staff quality is an important route to success. Still, it is important to note that all of these studies, bar one, were conducted in the United States in single purpose governments (school districts), and drew on largely archival measures. All of the studies, except that of Carmeli (2006), included measures of effectiveness. Until this set of studies is supplemented by more research in other settings, it is difficult to generalize about the benefits of staff quality.

Resource-based theories stress that personnel turnover can lead to declines in performance as organizations lose the vital expertise offered by experienced staff (Dess and Shaw 2001). Table 6 presents the results of our support score analysis for the effects of personnel stability on local government performance. Changes in personnel can be especially harmful to governmental performance because of the loss of investment in skills it represents and the costs of training new members of staff to replace those with experience and know-how (O'Toole and Meier 2003). As with staff quality, this aspect of local government management has largely been examined in special purpose governments in the United States, with just two studies conducted in the United Kingdom and to date none elsewhere. Over 50% (unweighted 52, weighted 52) of these studies support the hypothesis that personnel stability is a positive force in public organizations. One especially interesting feature of the evidence in these studies is that stability has a positive effect on performance across two different organizational echelons: the front-line; and senior managers. In all of these studies, stability was measured as staff turnover, but used a mix of archival and perceptual data. Effectiveness was investigated in each study, with equity and service quality also a focus in four and two studies, respectively.

The theory of representative bureaucracy suggests that public organizations are more responsive to service users' needs when their workforces reflect the demographic characteristics of those users. From a resource-based perspective, workforce representativeness therefore constitutes a key human resource to be harnessed by managers seeking to improve results. All but one of the 14 studies of representative bureaucracy that we analyzed were conducted in the United States, again in single purpose organizations, and drew exclusively on archival measures of management and performance (table 7). Ten of the studies included measures of effectiveness, and equity was examined in seven. Representative bureaucracy was operationalized at the organizational level and also across two echelons: the front-line and managers or administrators. The unweighted (61%) and weighted (43%) support score offer moderate legitimacy of the following thesis: that better representation of women and ethnic minorities in local governments appears, on average, to lead to better outcomes for those segments of society.¹³

¹³ The large number of tests in O'Toole and Meier (2003) skew the means, and removing this study results in strong support: the unweighted score is 63 and the weighted 52%.

Table 5
Staff Quality

Study	Sample	Country	Purpose	Management		Performance		No. tests	+	-	ns
				Staff Quality	Measure	Dimension	Measure				
Ruggiero et al. (1995)	636	United States	SP	Managerial quality	Archival	Eft	Archival	2	50	0	50
Meier et al. (1999)	527	United States	SP	Managerial quality	Archival	Eft, Eq	Archival	2	0	0	100
Meier and O'Toole (2001)	2,535	United States	SP	Managerial quality	Archival	Eft, Eq, Q	Archival	3	100	0	0
Meier and O'Toole (2002)	5,126	United States	SP	Managerial quality	Archival	Eft, Eq, Q	Archival	11	82	9	9
O'Toole and Meier (2003)	2,525	United States	SP	Managerial quality	Archival	Eft, Eq, Q	Archival	10	60	10	30
Carmeli (2006)	263	Israel	MP	TMT skills	Perceptual	Efy, In	Archival	4	100	0	0
O'Toole and Meier (2006)	4,114	United States	SP	Managerial quality	Archival	Eft, Eq, Q	Archival	1	100	0	0
Meier et al. (2007)	3,041	United States	SP	Managerial quality	Archival	Eft	Archival	9	78	11	11
Meier and O'Toole (2008)	4,114	United States	SP	Managerial quality	Archival	Eft	Archival	1	100	0	0
O'Toole and Meier (2009)	2,400	United States	SP	Human capital	Perceptual	Eft, Eq, Q	Archival	16	88	6	6
Pitts and Jarry (2009)	1,041	United States	SP	Teacher human capital	Archival	Eft	Archival	1	0	0	100
Meier and O'Toole (2010)	4,114	United States	SP	Managerial quality	Archival	Eft	Archival	6	100	0	0
Meier et al. (2010)	3,041	United States	SP	Managerial quality	Archival	Eft	Archival	2	100	0	0
Johansen (2012)	4,664	United States	SP	Managerial quality	Archival	Eft, Eq	Archival	13	68	32	0
						Number of studies		14			
						Number of tests		81			
						Unweighted			73	3	25
						Weighted			77	5	18

Note: Purpose: MP = multipurpose, SP = single purpose; Performance dimension: Eft = Effectiveness, Efy = Efficiency, Eq = Equity, In = Index Q = Quality, TMT = top management team.

Table 6
Personnel Stability

Study	Sample	Country	Purpose	Management			Performance			No. of tests		
				Stability	Measure	Dimension	Measure	Dimension	+	-	ns	
Davies and Coles (1981)	41	United Kingdom	SP	Personnel	Archival	Eft	Archival	2	100	0	0	0
O'Toole and Meier (2003)	2,525	United States	SP	Teacher & Managerial	Archival/Perceptual	Eft, Eq, Q	Archival	20	65	10	25	25
Bohte (2004)	1,046	United States	SP	Teacher	Archival	Eft	Archival	5	0	100	0	0
Meier et al. (2006)	1,039	United States	SP	Teacher	Archival	Eft, Eq	Archival	5	40	60	0	0
O'Toole and Meier (2006)	4,114	United States	SP	Teacher & Managerial	Archival/Perceptual	Eft, Eq	Archival	2	100	0	0	0
Meier et al. (2007)	3,041	United States	SP	Teacher & Managerial	Archival/Perceptual	Eft, Eq, Q	Archival	18	67	0	33	33
Meier and Hicklin (2008)	4,315	United States	SP	Turnover	Archival/Perceptual	Eft	Archival	6	0	50	50	50
Meier and O'Toole (2008)	4,114	United States	SP	Teacher & Managerial	Archival/Perceptual	Eft	Archival	2	100	0	0	0
Pitts and Jarry (2009)	1,041	United States	SP	Teacher	Archival	Eft	Archival	1	100	0	0	0
O'Toole and Meier (2009)	2,400	United States	SP	Teacher & Managerial	Archival/Perceptual	Eft	Archival	2	50	50	0	0
Meier and O'Toole (2010)	4,114	United States	SP	Teacher & Managerial	Archival/Perceptual	Eft	Archival	12	67	0	33	33
Meier et al. (2010)	3,041	United States	SP	Teacher	Archival	Eft	Archival	4	50	0	50	50
Meier, O'Toole, and Hicklin (2010)	703	United States	SP	Teacher	Archival	Eft	Archival	3	0	0	100	100
Boyne et al. (2011a)	148	United Kingdom	MP	TMT	Archival	In	Both	4	0	0	100	100
						Number of studies		14				
						Number of tests		86				
						Unweighted		53	16	32		
						Weighted		52	15	32		

Note: Purpose: MP = multipurpose, SP = single purpose; Performance dimension: Eft = Effectiveness, Efy = Efficiency, Eq = Equity, In = Index Q = Quality.

Table 7
Representative Bureaucracy

Study	Sample	Country	Purpose	Management		Performance					
				Representative Bureaucracy	Measure	Dimension	Measure	No. of tests	+ - ns		
Meier and Stewart (1992)	67	United States	SP	Black teacher & manager	Archival	Eft, Eq	Archival	14	43	14	43
Meier (1993)	132	United States	SP	Latino teacher & manager	Archival	Eft, Eq	Archival	4	50	25	25
Meier et al. (1999)	527	United States	SP	Minority teachers	Archival	Eft, Eq	Archival	2	50	0	50
Meier and Bothe (2001)	2,860	United States	SP	Minority teachers	Archival	Eq	Archival	1	100	0	0
Meier and O'Toole (2003)	1,043	United States	SP	Latino administrator & manager	Archival	Eft, Eq	Archival	54	31	0	69
Smith (2003)	179	United States	SP	Black representation	Archival	Q	Archival	3	100	0	0
Andrews et al. (2005)	150	United Kingdom	MP	Ethnic diversity senior management team	Archival	In, US	Archival	4	0	75	25
Pitts (2005)	2,482	United States	SP	Teacher & manager diversity	Archival	Eft, Q	Archival	12	33	17	50
Meier and Nicholson-Crotty (2006)	60	United States	SP	Female police	Archival	Eft	Archival	6	100	0	0
Meier et al. (2006)	1,039	United States	SP	Female teachers	Archival	Eft, Eq	Archival	5	80	0	20
Pitts (2007)	6,994	United States	SP	Teacher & manager diversity	Archival	Eft, Eq, Q	Archival	12	33	17	50
Wilkins and Williams (2008)	96	United States	SP	Black representation	Archival	Eq	Archival	1	100	0	0
Pitts and Jarry (2009)	1,041	United States	SP	Teacher diversity	Archival	Eft	Archival	3	67	33	0
Roch and Pitts (2012)	1,263	United States	SP	Teacher & manager diversity	Archival	Eft	Archival	8	63	0	37
						Number of studies		14			
						Number of tests		129			
						Unweighted		61	13	26	
						Weighted		43	9	48	

Note: Purpose: MP = multipurpose, SP = single purpose; Performance dimension: Eft = Effectiveness, Efy = Efficiency, Eq = Equity, Q = Quality, US = User satisfaction.

- Theobald, Nick A., and Sean Nicholson-Crotty. 2005. The many faces of span of control: Organizational structure across multiple goals. *Administration and Society* 36:648–60.*
- Thompson, James D. 1967. *Organizations in action*. New York, NY: McGraw-Hill.
- Vinzant, Janet C., and Lane Crothers. 1998. *Street-Level leadership: Discretion and legitimacy in front-line public service*. Washington, DC: Georgetown Univ. Press.
- Walker, Richard M., and George A. Boyne. 2006. Public management reform and organizational performance: An empirical assessment of the UK Labour government's public service improvement strategy. *Journal of Policy Analysis and Management* 25:371–93.*
- Walker, Richard M., Boyne, George A., and Gene A. Brewer (eds.). 2010. *Public management and performance: Research directions*. Cambridge: Cambridge Univ. Press.
- Walker, Richard M., and Gene A. Brewer. 2009. Can management strategy minimize the impact of red tape on organizational performance? *Administration and Society* 41:234–48.*
- Walker, Richard M., Damanpour, Fariborz, and Carlos Alberto Devece. 2011. Management innovation and organizational performance: The mediating effect of performance management. *Journal of Public Administration Research and Theory* 21:367–86.*
- Walker, Richard M., Andrews, Rhys, Boyne, George A., Meier, Kenneth J., and Laurence J. O'Toole, Jr. 2010. Wake up call: Strategic management, network alarms and performance. *Public Administration Review* 70:731–41.*
- Walker, Richard M., Brewer, Gene A., Boyne, George A., and Claudia N. Avellaneda. 2011. Market orientation and public service performance: NPM gone mad? *Public Administration Review* 71:707–17.*
- Walker, Robert, and Julie Williams. 1986. Housing benefit: Some determinants of administrative performance. *Policy & Politics* 14:309–34.*
- Wilkins, Vicky, and Brian Williams. 2008. Black or blue: Racial profiling and representative bureaucracy. *Public Administration Review* 68:654–64.*
- Williams, Daniel W. 2003. Measuring government in the early Twentieth Century. *Public Administration Review* 63:643–59.