Measuring Government in the Early Twentieth Century

This article discusses the early history of performance and productivity measurement. It finds sophisticated development of these tools beginning in the first decade of the twentieth century, primarily at the New York Bureau of Municipal Research. These practices grew out of accounting, the social survey, work records, and municipal statistics. The bureau built government's capacity to measure. They advocated such basic empirical practices as making observations at all, doing so systematically and routinely, and recording data at the time of observation. By 1912, performance measurement exhibited many of the features associated with the modern practice: measuring of input, output, and results; attempting to make government more productive; making reports comparable among communities; and focusing on allocation and accountability. Performance measurement was developed in the context of shifting power between the elected executive and the legislature.

On October 25, 2000, the president of the New York City Citizens Budget Commission described a difficulty with obtaining performance data from the New York City Police Department. She concluded by saying that if the police claimed they did not have the data, there was little that could be done. Back in 1906, the New York Bureau of City Betterment found itself similarly situated by the reluctant Manhattan borough president John F. Ahearn. The bureau chose a different tack: “[S]omeone suggested they didn’t need the record for everything. The city had to invite bids for contracts and had to issue specifications in some detail. Then the Bureau had only to see whether the city was spending according to contract specifications. Among the things that the staff did was to stand around counting sacks of cement” (Dahlberg 1966, 27–28, note 23).

As a consequence, Ahearn sued the bureau for libel, he was later removed for incompetence, and the practice now known as performance measurement was born.

Any modern text on performance measurement—and there are many (Ammons 1995, 2001a; Berman 1998; Hatry 1999; Hatry et al. 1999; Heinrich and Lynn 2000; Kearney and Berman 1999; Sims 1998; Williams, Webb, and Phillips 1991)—provides advice on such matters as inputs, outputs, outcomes, and benchmarks. The general thesis of these books is that government needs to observe itself to see whether and how it meets public objectives, how it uses resources to do so, and how its processes can be made more productive. One is advised to avoid measuring inputs alone, or just inputs and outputs. Emphasis should be on outcomes, citizen satisfaction, and like matters. Benchmarks should be set based on the best performance observed among other organizations. Productivity should be improved over time. Sometimes performance measurement is suggested as a productivity tool, aimed at obtaining the greatest possible effect with available resources. At other times, the emphasis is on accountability to the public, making sure that government is truly pursuing the will of the public.

During the 1990s, performance measurement, productivity improvement, and performance budgeting were part of such agendas as Vice President Gore’s National Performance Review; the Government Performance and Results Act; the reinventing government literature (Osborne and Gaebler 1993; Osborne and Plastrik 1998), which is linked to the National Performance Review and to New Public
Management; and local government accounting guidelines issued by the Government Accounting Standards Board (GASB 2001). Performance measurement has been promoted before, in the early 1970s, responding to President Nixon’s and Senator Proxmire’s concerns, and in 1949 as a recommendation of the first Hoover Commission.

However, the early history of performance measurement is poorly known. Geert Bouckaert’s overview of the history of productivity improvement contains only five citations from before 1920 (Bouckaert 1990, 1992), and, of these, four are only tangentially related to performance measurement. Relying on secondary sources that primarily address other matters, he asserts the well-worn view that these early innovators were concerned exclusively with reducing the cost of government services, and he explicitly denies they addressed outcomes (Bouckaert 1992, 17). I will provide contrary evidence later in this article. Other works addressing performance-measurement history provide even more cursory treatment of this period (Balk 1992; Miller 1990; Rich and Winn 1992; White 1999; Lynch and Day 1996; Nyhan and Marlow 1995) or are focused elsewhere (Arnold 1986; 1995; Blunt 1989; Conant 1986; Golembiewski 1989; Martin 1990). One must turn to general histories (Mosher 1968), to specific accounts of the bureaus of municipal research (Dahlberg 1966; Gill 1944; Haber 1964; Schachter 1997; Stivers 2000), or to the origins of the Brookings Institute (Crichtlow 1985; Kahn 1997) to find more in-depth discussion, and these do not point out the significance of the performance- and productivity-measurement activities.

This article examines the origin of performance- and productivity-measurement practices. By “origin,” I mean the earliest identifiable sustained effort to engage in performance measurement and productivity improvement. This sustained effort begins with the measurement practices of the New York Bureau of Municipal Research (Crane, Fairlie, and Merriam 1923; Merriam 1923; Woodruff 1910), although other partial efforts originated in the same period or even somewhat earlier. This article focuses on the early years of the bureau, up to 1912.3

The New York Bureau of Municipal Research and the Bureaus

It is well-known that the New York Bureau of Municipal Research and associated organizations were instrumental in promoting the applied study of public administration in its formative years. These developments are often characterized as the origins of modern budgeting (Dahlberg 1966; Kahn 1997; Schachter 1997; Stivers 2000). What is less well-known is that a large part of what the bureau did was to perform and promote the measurement of government performance and productivity.

A thorough examination of the conditions leading to the bureau’s performance-measurement developments is outside the scope of this article, but a brief review sets the context of this development.4 Following a discussion of the bureau’s techniques, I will discuss three sources in more detail. The New York Bureau of Municipal Research was incorporated in 1907 following a year of temporary existence as the Bureau of City Betterment. During that first year, the bureau was headed by Henry Bruere. Beginning with incorporation, it was headed by William H. Allen, Henry Bruere, and Frederick A. Cleveland. All three co-directors had prior experience with social welfare agencies: Allen and Bruere as employees, Cleveland as an accountant advisor. For a variety of reasons, Allen, Bruere, and Cleveland can be linked with prior statistical science practices, including the survey method associated with the settlement houses of that era, municipal accounting practices, and Frederick Taylor’s scientific management. They also can be linked to the Progressive reform movement, which was rapidly developing a theory of executive government with legislative oversight as part of its new focus on competence, rather than mere goodness, in government officers. All of these links contributed to setting the bureau’s agenda. A more extensive account of the bureau can be found in the work of Crichtlow (1985), Dahlberg (1966), Gill (1944), Kahn (1997), Haber (1964), Schachter (1989), Stivers (2000), and Waldo (1948).

At the end of the nineteenth century, cities became important government units, largely because they contained a large and growing proportion of the population. At the same time, their governments were widely viewed as corrupt. After the Civil War, municipal reform was a recurrent theme in popular and academic discussions. By the 1890s, it had become apparent that it was not enough to throw the rascals out. Reformers shifted their focus from moral reform of government officers to reform of government processes. Emblematic, perhaps instrumental, in this shift was the formation of the National Municipal League in 1894 (Woodruff 1908a).

A tenet of the municipal reform program was that governments could be made more accountable by subordinating all administrative functions under a single person—the governmental executive—who could be made personally responsible for government (Low 1889; MacVeagh 1896). In part, it was thought that diffuse responsibility through the legislature made each individual election too unimportant to attract responsible voter attention. The governmental executive, on the other hand, would be important enough to get that attention. It was also thought that legislators had little motive to control government expense, and a greater interest in spending money in ways that specifically benefited their constituencies (Bradford 1893). Cooperation between legislators was obtained through log-
The executive would be concerned about the interest of the whole jurisdiction, and therefore would be less inclined to permit expensive projects that benefited narrow constituencies (Flack 1910). For example, during the late 1800s, New York City charter reforms continually narrowed the discretion of the city council in favor of mayoral power (Durand 1900; Godkin 1894); the formation of the Board of Estimate limited the city council’s role in budget making to not much more than a plebiscite (Ivins 1902). The other involved a consolidation of the executive branch of government by eliminating numerous minor elective administrative positions (Goodnow 1902).

The role of the responsible executive changed over time from a morally responsible executive, characterized, at least by some, as coming from the upper classes of society, to a competent executive, modeled on the executive of a large business. After the disappointing administration of Seth Low in New York from 1902 to 1903, the view shifted again, from competent executive to a competent organization (Burks 1912b).

It would be a mistake, however, to think there was universal consensus supporting the singular-executive model. The short-lived promotion of commission government moved in the opposite direction, shifting executive power into the hands of the legislature. The bureau was actively opposed to commission government (Bruere 1912b); however, the bureau’s codirectors’ views concerning the roles of executive and legislature were in conflict and became the subject of a heated debate several years after Allen left the bureau (Allen 1917; Chase 1917). Those who offered alternatives to the singular-executive model included E. Dana Durand, John Commons, and Clinton Woodruff, who was the secretary of the National Municipal League (Commons 1902; Durand 1900; Woodruff 1903). Performance measurement and reporting played a role in reconciling these conflicting views. Through the instrument of reporting, the executive could be given broad discretion while retaining the more inclusive democracy of legislative decision making, or at least its appearance.

Many of these themes are familiar to us today. Promotion of the line-item veto in the early 1990s reflected a continued belief that the singular executive represents the public as a whole and will act more responsibly than pork-barreling, logrolling legislators. The New Public Management approach emphasizes an exchange of discretion to administrators for accountability through reports to elective officials. Taxpayer concerns frequently influence American politics at all levels. And since the 1970s, popular rhetoric has hammered at public employees with an accusation of incompetence, which, it is argued, must be supplanted with business practices.

At the beginning of the twentieth century, the bureau and other bureaus that emulated it also sought to improve government by bringing business practices into government. They focused much of their effort on improving the way government does its work, that is, improving productivity and accountability. This effort had two major components: The first was gathering data through accounting, record keeping, and needs assessment. The second was converting the data into useful information and using it through reporting, budgeting, and productivity improvement. Today these developments may seem trivial, but during the first decade of the twentieth century these developments were the cutting edge of government reform, and they form the conceptual foundation for performance measurement. By understanding the elements of this empirical approach to government management, we may develop a better understanding of these measurement practices.

### Gathering Data

Through extensive attention to observation and record keeping, the bureau introduced the systematic and routine collection of original data defined for the purpose of observing government at work (Merriam 1923). Prior to this time, accounting served more basic money management purposes. In the pre-bureau era, data were sometimes gathered for policy making, but on an ad hoc basis. No system existed for either needs assessment or outcomes monitoring. The bureau established a program of collecting data as the first step of performance measurement. Precise measurement replaced vague impressions to inform decision making and to verify claims of performance.

### Accounting

Accounting, record keeping, and needs assessment formed the core of this effort by producing the empirical basis for reporting, budgeting, and productivity improvement. In the May 1912 edition of the *Annals of the American Academy of Political and Social Science*, which was titled “Efficiency in City Government,” four articles addressed accounting and a fifth addressed revenue collections by city utilities (Clowes 1912; Fisher 1912; Prendergast 1912; Taussig 1912; Walton 1912). Accounting is also addressed in Bruere’s *The New City Government* (1912b). Frederick A. Cleveland came to the bureau after teaching accounting at New York University and wrote a textbook on governmental accounting (Cleve-
The bureau influenced the Metz Fund to sponsor a series of talks on accounting, the text of which was sent to 325 municipalities, and compiled these talks into the *Handbook of Municipal Accounting* (BMR 1919 [1913]). The bureau focused on classifying data into functions so that costs could be determined and results could be compared with costs (Lapp 1909; Prendergast 1912; Sands and Lindars 1912). Many other accounting concepts introduced or promoted by the bureau were needed for determining costs, the point of which was to associate these costs with results and determine administrative efficiency (Cleveland 1904; Fisher 1912; Goodrich and Holton 1912; Walton 1912).

**Record Keeping**

Improved accounting depended on effective record Keeping: “These preliminary studies showed that department heads were handicapped in any effort to improve their departments by the lack of proper records and account ...” (Treleven 1912, 273). In addition to financial records, there was a focus on work records and records of outputs, outcomes, social indicators, and needs.

**Work Records.** Each form of record keeping required the development or implementation of measurement techniques. For work records, the bureaus recommended the use of time sheets and work plans (Allen 1908; Clowes 1912; Goodrich and Holton 1912; Hopkins 1912; Pultz 1912). Benjamin F. Welton argued, “The most common, and at the same time, most serious loss of labor efficiency is due to waste of time.... The lack of accurate work records [prevents comparison with work standards].... More often than might be expected, the efficient performance of work is subordinated to accounting convenience” (Welton 1912, 110). Henry Bruere said, “Efficiency records must be specially devised for each class of employment, and must be based upon comprehensive understanding of the positions to which they relate.” He described New York City’s efficiency records as containing “Quality of work performed [,] Quantity of work performed [,.] General conduct [,.] Punctuality and attendance [,. and] Executive ability and capacity for initiative where work is of a character that will permit definite determination.” (Bruere 1912b, 345) However, he objected that these records were based on retrospective estimates rather than more timely data collection.

**Output.** The bureaus recognized the need for output data—for example, Fritz Reichmann expressed a need for “a full, lucid report on work which has been done ...” (Reichmann 1912, 217). Early work of the New York Bureau of Municipal Research focused on determining whether outputs existed at all (Braddock 1912; Schchter 1989; Welton 1912). An unattributed article on the report of the Commission on Economy and Efficiency (1911, 627–28) reports, “Its most important feature is that it calls for ... [reporting] in such a way that expenditures will be shown in terms of ... services operated and work performed...” (italics added). Yet, this literature provides little insight into how output was observed or recorded, although Allen does ask whether the right outputs were being measured (Allen 1909).

**Outcomes.** The bureaus gave considerable attention to the measurement of outcomes. A few years earlier, in 1896, Walter F. Wilcox advised, “[T]he benefit of a sewerage system should be measured in terms of decreased mortality rather than in terms of increased productivity” (378). Wilcox plainly argued that government should be held accountable primarily for outcomes, not merely the conservation of government funds. In 1912, Frank P. Bachman examined measurement outcomes in education, where the intended effects are delayed in time 5–10 years, so “other criteria” are needed. He recommends interim indicators such as attendance, graduation, and success on tests. He discusses effectiveness in the form of improved test results over time. He emphasizes the need to link efficiency standards to schools that measure up on these factors, not to the least expensive schools (Bachman 1912). Thus, for him efficiency is not mere economy, it is a combination of high quality at relatively low cost, that is, optimal effective use of resources.

In Joseph F. Neff’s “Efficiency In Child Saving,” outcomes are related to reduced infant mortality. Neff provides guidance for paying particular attention to those districts where infant mortality is severe in the first place (Neff 1912).

William H. Allen’s *Efficient Democracy* is a handbook on the development of outcome measures. He says, “Efficiency asks that the new ideals prove their superiority by a show of results” (Allen 1907, 193, italics in original). Following this assertion he provides a list of likely measures such as “probationers re-arrested + total number of probationers” and “percentage of probationers re-arrested + percentage paroled without supervision re-arrested.” Allen’s outcome measures are adjusted through ratios to provide analytic insight into the effects of policies.

This interest in outcomes reflects a sophistication in methods and some understanding of the difficulty in establishing a causal link between program activity and outcomes. However, the development of inferential methods for determining causality was still in early development in the first decade of the twentieth century, so, understandably, the bureaus treated this matter rather simply. For example, today, Allen’s re-arrest ratio would be the object of analysis rather than its result.

**Social Indicators and Needs Assessment.** Beyond outcome measures, the bureaus pursued a study of social indicators and policy problems. In the narrowest sense, these...
social indicators are similar to outcomes measures. Thus, Neff’s infant mortality measures could fall within this classification. However, following the practices of the settlement houses (Stivers 2000), these surveys were often general, focusing on matters at or beyond the current margins of public policies. For example, Bruere recommends collecting such health statistics as “Death rate; Increase or diminution in tuberculosis; Number cases typhoid; Infant mortality rate; Measles, scarlet fever and diphtheria rates (morbidity and death); Bacteria count of milk—maximum, minimum and average; Number of school children treated for defects; Number of nuisances abated, etc., etc.” (Bruere 1912b, 122).

Beyond the mere collection of social-indicator data, the bureaus pursued full-scale policy analysis: “The social survey is based on the theory that one measure of the efficiency of government is the extent to which it promotes the welfare of the citizens. Since year by year the scope of the activities of the municipality broadens to include new fields of social service, there is a constantly increasing need of careful study to determine how these social problems can best be solved” (Treleven 1912, 272).

The sorts of problems studied by just one bureau included child welfare, housing, infant mortality, legal aid, “woman’s wages,” and garnishment.

**Useful Information**

Data alone were of little value. The objective of gathering empirical data was to communicate what the government was doing and to inform decision making. Once data were collected, they were compiled in reports to higher-ranking administrators, political authorities, and the public. The reports served political and managerial purposes and contributed to informed decision making and better management.

**Reporting**

In 1904, Frederick A. Cleveland said, “It is coming to be understood that we need, not constitutional and legal restrictions placed on official conduct, but measures which will provide for greater efficiency, and which will secure a better understanding of municipal activity and its results. The most important of these measures are those which aim to provide a more enlightened system of official records and reports” (1904, 397). While Cleveland was primarily discussing New York City, the problem was general. Frederick R. Clow quotes Nathan Matthews, Jr., mayor of Boston from 1891 to 1895: “[B]elieving that the first duty of the public officer charged with the disbursement of millions of dollars of the public money was to search the printed reports of the city government for accounts that would show the cost, from year to year, of equipping and maintaining the various departments of municipal services, I was amazed to discover that practically there were none” (Clow 1896, 460–61, citing “Valedictory Address,” 20–21).

Clow continues, “What must be the difficulty of the average citizen of the average American city in trying to understand his city’s finances?” (1896, 461).

One point of reporting was to show the link between the money spent by government and the services provided: “[I]t will be necessary first to establish records which develop facts descriptive of the work of the department, and by means of which its expenditures may be contrasted with results” (Bruere 1908, 114–15). Reporting, thus, was intimately associated with accounting. However, reports were not simply accounting summaries. Reports were expected to provide “intelligence” (information) necessary for managing the organization: “Some provision must be made not only for procuring a type of intelligence in management competent to exercise the sort of discretion required, but also for a complete and effective means of obtaining information necessary to such discretion in the exercise of administrative control by the officers having the business in hand” (Cleveland 1909, 208–9).

The sort of information to be included in the report included need, unit costs, results, and gains and losses (Allen 1908, 1912; Baldwin 1908; Bruere 1908, 1912a; Cartwright 1912; Halbert 1912; Pultz 1912; Welton 1912). There was considerable concern that such reports should assign costs narrowly to functions so they could be comparable across time, jurisdictions, organizations, and work units (Allen 1906; Bachman 1912; Cleveland 1912a; Lapp 1909; Merriam 1912; Pearse et al. 1899; Pultz 1912; Taussig 1912; Walton 1912).

A unique form of reporting innovated by the New York Bureau of Municipal Research was called the “Budget Exhibit,” an event conducted in a public hall over an extended period. It was sponsored by the bureau in 1908 and 1909, then by New York City in 1910 and 1911 (Allen 1908; Beard 1970; Braddock 1912; Sands and Lindars 1912). The exhibit was a tangible performance report: “Facts and figures graphically displayed, intermingled with physical objects, informed the visitor of the city’s activities—what had been and what was expected to be done with the taxpayer’s money” (Sands and Lindars 1912, 148).

These exhibits were part of the bureau’s publicity program. Publicity was highly valued by the bureaus (Allen 1908; Bruere 1908, 1912a; Burks 1912b; Cleveland 1904, 1912b; Kahn 1997; Miles 1912; Schachter 1997). The objective of publicity was to make “efficient citizens.” Bruere enumerates five standards of government efficiency, including “[e]fficient citizenship, equipped through intelligence regarding citizen needs, and armed with facts regarding government conditions and results, to co-operate with the
public officials in enlarging the usefulness of government and insuring its adherence to established standards of efficiency” (Bruere 1912a, 76). Citizens needed information to fulfill their duty to keep watch on government. This need to communicate with the public permeates the whole of the early bureaus’ literature. The bureaus saw themselves as organs of democracy, reforming government through research that provided methods of accomplishing the complex goals of government and eliminating waste and dishonesty while retaining democratic control over objectives. The budget exhibits were used to hold public managers accountable for their use of public funds (Braddock 1912).

The Budget

Having linked work, output, outcomes, social indicators, and needs with actual expenditures for the past and current periods, the next step was to link all of these to proposals for expenditures and programs—that is, the budget. Henry Bruere says, “Reluctance to appropriate has been partly attributable to the uncertainty that funds requested would be used with fullest efficiency. But it has also been due to the fact that neither the requesting departments, the community, nor the board of estimate and apportionment, had definitely pictured the wide gap existing between health and charitable services now rendered and the health and charitable needs of the city” (Bruere 1912a, 10).

The bureau developed the “segregated” or functional budget, which separated expenditures by major activity or function. The functional budget was designed to clarify discretion and provide for accountability. Accountability came in the form of reports that showed expenditures within functions. These reports could be coordinated with past and future budgets, needs and accomplishments, and work productivity. Thus, functional budgeting and accounting formed the nexus of the rational management system and made it possible to base the budget request on empirical data (Prendergast 1912).

The functional categories allowed comparison of funding requests between similar work units, similar functions in different jurisdictions, and funding in the same function across time. Thus, these categories allowed for challenging a budget request based on excessive costs and for assigning an activity to a more efficient work unit.

Productivity Improvement

Henry Bruere writes of the “[T]ask, necessarily first undertaken, of instituting efficient business practices in city government” (Bruere 1912a, 4). These business practices were to associate costs with results and to stamp out waste. The three primary tools for stamping out waste were comparison, records, and method.

Comparison was achieved through functional budgeting and accounting, which also served as the cost accounting system. As projections and expenditures were classified by functions, sub-functions, and work units, costs could be compared across similar functions and across the same function in various localities. This comparison helped to find work units that were inefficient or extravagant. Efficiency—or its nemesis, inefficiency—was determined by the efficiency ratio, units of something desired divided by units of resources or vice versa (Allen 1907; Prendergast 1912). When the ratio showed unusually high resource consumption for a unit desideratum, there was prima facie evidence of waste, which must then be examined in more detail. However, the objective was not merely to drive down costs. Cost benchmarks were pegged to quality outcome, not the cheapest service delivery (Bachman 1912). This view has not changed over the past century, although it has become more apparent that efficiency and quality are hard to capture in a single measurement (Bouckaert 1995; Holzer 1991; Kravchuk and Schack 1996; Nyhan and Marlow 1995).

Records were a second implement for stamping out waste. As we have already seen, records were a necessary component of accounting, reporting, and budgeting. However, they also served a direct role in productivity by exercising control over contracting, purchasing, and work. As Charles E. Merriam observed, “The amount of work done in the various departments materially increased under the stimulus of inquiry” (1912, 297).

While wasteful financial transactions were an object of concern (Cleveland 1909; Miles 1912; Pultz 1912), the main focus was on inefficient work (Burks 1912b; Welton 1912; Woodruff 1908b). J. Leggett Pultz advised, “One of the most important features of efficiency is to promote the personal efficiency of the individual employee. This is accomplished by means of service records which show the degree of proficiency in the performance of his duties, allows the acknowledgement of efficient service, discloses inefficiency, and places responsibility” (1912, 81–82). The role of work records was to get workers to work: “To measure the efficiency of services, adequate time reports with proper certification and approval are required. Forms of this kind put a check upon one of the greatest causes of waste and set up standards of labor and efficiency which automatically separate the workers from those who cannot or will not perform the service due, demanded and paid for” (Taussig 1912, 60).

Benjamin F. Welton determined, “[M]unicipalities normally suffered a loss of efficiency ... varying from 40 per cent to 70 per cent” (1912, 103–104). Cleveland equated this loss with “theft of time and services” (1909, 26). Work plans, time sheets, and observation were recommended.

In addition to improving productivity through observation, the bureaus recommended improvement in method. First, the bureaus recommended work plans (Clowes 1912; Dunaway 1916; Welton 1912), which set clear expecta-
tions of accomplishment over a limited period of time. This arrangement produced verifiability: A supervisor could confirm that the planned work actually occurred. Work plans also allowed planners to organize work to minimize waste. Thus, the work plan served to improve the work that municipal workers were already doing.

Work planning gradually gave way to standardization (Burks 1912a) and the use of labor saving or time saving devices (Klein 1912). Standardization can be understood in two senses. First, work was made uniform through such devices as written guidelines for work performance (Clowes 1912). Second, specific expectations were set for what counted as doing the job (Bruere 1912a; Welton 1912). These forms of productivity improvement went beyond the systematic and verifiable performance of the usual work; the work itself was redefined to be more efficient. These steps reflected either a new or broadening influence of scientific management after 1910.

**Summary of Measurement Activities**

The bureau advocated systematic and routine observation of government by collecting data on inputs, outputs, outcomes, and needs. These data were used to communicate accountability information to office holders and the public, and to inform budget making and productivity improvement. Through measurement, the bureau sought to substitute facts for vague impressions. Facts served as a check against error when holding administrators accountable, allocating resources, or improving work processes. To make this possible, the bureau advanced an information system based on the functional accounting scheme. This scheme served not only as an accounting system, but also as a method of cost accounting and as the management information system.

**Sources of the Bureau’s Techniques**

Frederick Thayer has said that modern performance measurement and productivity improvement are merely a reiteration of scientific management (Thayer 1972). If so, the link should arise with the activities reviewed here. However, the evidence points elsewhere. These performance-measurement activities began early with the bureau, while it still operated as the Bureau of City Betterment in 1906. While the citations in this article are weighted to the later part of the seven-year period under discussion, this result is mostly an incidental result of the focus on the bureau’s approach in the May 1912 issue of the *Annals of the American Academy of Political and Social Sciences*. The links between the bureau and scientific management are fairly weak in the early part of this period, when these activities arose. While Dahlberg suggests a link as early as 1903 through Allen’s exposure to Frederick Taylor’s *Shop Management* (Dahlberg 1966; Taylor 1947), his inspiration through this source was not so strong that it was shared with Cleveland. It was the wide popularization of scientific management through the Eastern Rate case of 1910–11, which brought scientific management to Cleveland’s attention (Haber 1964). The strong link between the bureau and scientific management is generally understood to derive from Cleveland’s interests. Cleveland became an active member of the Taylor society, and, during the second decade of the twentieth century, the bureau became more aligned with scientific management. Little of this link occurred during the early period under discussion here. The strongest link is found in the transition between early and later productivity-improvement processes. In the early period, the bureau advocated getting workers to do the work they were hired to do. While some of these early techniques, such as keeping detailed work records, appear to be influenced by scientific management, the core scientific management concept—time studies—is completely absent. After 1910, however, a second theme was added, a call for systematic work planning, development of work standards and standardization, and the use of modern equipment. This last-named theme reflects the more direct influence of scientific management. The most direct links, however, arose even later, particularly under the influence of Morris Cooke (Schachter 1989).

If not scientific management, where did these techniques originate? As suggested in endnote 4, there are many sources of these techniques. However, two stand out: First is the survey method of the settlement houses. Through their involvement in the Association for the Improving the Condition of the Poor (AICP), all three bureau directors were closely linked to the social work community, where settlement houses played a prominent role—the AICP later opened a settlement house—and Bruere worked for settlement houses before working for the bureau (Dahlberg 1966; Kahn 1997; Kraus 1980; Stivers 2000). During the late 1800s, the settlement houses imported and made their own a method of research originated in England by Charles Booth (Bales 1991; Blumer, Bales, and Sklar 1991; Converse 1987; Sklar 1991; Stivers 2000). This method, known as the survey or the social survey, was a comprehensive investigation of the conditions within a small community. The word “survey” is not the modern word which implies sampling, questionnaires, and use of inferential statistics, but its precursor, the survey of a situation, which is taking a good look at matters. Booth’s surveys looked more like a census, although of a very small area and, as adopted by the settlement houses, provided a comprehensive look at the conditions of the poor. The use of “survey” language was carried over into the literature of the bureau, where they considered their core technique, particularly as they served new communities, to be the survey. The bureau’s surveys were also comprehensive.

"Measuring Government in the Early Twentieth Century" 649
However, the bureau applied them to the conditions of government: the political environment, the workings of government, citizen satisfaction, and outcomes. The bureau continued to describe their empirical method as the survey well beyond the time discussed here and conducted surveys for governments both within the United States and internationally well into the 1920s.

The link between the bureau’s approach and the social survey can be seen in the similarity of practices. For example, the bureau advocated taking photographs of streets for careful analysis of their condition (Bruere 1912b, 32). This same approach can be seen in a typical social survey from the era, such as the Social Survey of Portsmouth, Ohio (Mark 1916), which contains more than 20 photographs of street and housing conditions. This same survey contains several typical, detailed maps coded for social conditions following the style of the Booth and Hull House surveys (Blumer, Bales, and Sklar 1991; Sklar 1991). That this survey reflects the bureau’s practices is not surprising, as it was conducted by one of the many bureaus that emulated it. However, it also shows an integration of the bureau’s practices with those following the settlement house survey approach. The link between these practices was acknowledged and discussed during the early twentieth century (Bartlett 1928).

While the social survey accounts for the framework of the bureau’s empirical activities, we must look to accounting for the financial focus of its activities. Before coming to the bureau, Frederick Cleveland taught accounting at New York University and worked in the accounting firm of Haskins and Sells, where his mentor was Charles Waldo Haskins. Haskins was personally involved in the reform of governmental accounting in the late 1890s and early 1900s. The scope of this relationship and the depth of Haskins’s knowledge of governmental accounting reform is well documented in Jonathan Kahn’s Budgeting Democracy (Kahn 1997). Governmental budgeting and accounting were discussed in the newly emerging academic journals of the era, for example, in articles by Frederick Clow (1896, 1901). The close linkages of this community can be seen with a 1904 book review by Clow of a book on accounting written by Haskins with a foreword by Cleveland (Clow 1904). This link to the accounting profession during an era of strengthening cost accounting practices and specific focus on governmental accounting leads directly into the sort of practices discussed in earlier sections of this article.

The Bureau’s Program Questioned

We turn now to a consideration of an enduring criticism of the Bureau’s activities followed by a difficulty with the program of accountability.

The Pecuniary Issue

Dwight Waldo said of the bureau, “There have always been reformers … and taxpayers organizations, whose primary purpose is the lowering of taxes, no matter what the cost in human values. ‘Efficiency’ means for them, economy, ‘economy’ means less money spent by government, more retained by taxpayers—simply said” (1948, 195). On this view, the bureau’s achievement, if not their aim, was to provide a means of reducing the cost of government. The dividend of this reduction accrued to the taxpayer. Camilla Stivers (2000) recently reiterated this view. Others, such as Jane Dahlberg (1966), Jonathan Kahn (1997), and Hindy Lauer Schachter (1989, 1997), argue this view is inadequate, that the chief aim was to broaden the capacity of government to deliver services.

This issue is best understood in political context. On one view of the shifting power from legislature to executive, middle- and upper-income taxpayers engaged in the Mugwump reforms of the late 1800s and the Progressive reforms of the early 1900s to protect their tax dollars from the lower economic classes, who had gained the right to vote through progressive relaxation of property qualifications for the franchise during the mid-1800s. On this view, anticorruption and government professionalization reforms were aimed primarily at blocking the lower economic classes from engaging in redistributive tax and expenditure programs. This objective was certainly expressed during the late 1800s (Godkin 1894). Waldo’s accusation, more mildly asserted by Charles Merriam in 1923, was that the bureau was effectively a participant in this political agenda.

This debate arose during the period of the bureaus. Bruere said, “It was not a tax-saving incentive nor a desire for economy that inspired this first effort to apply modern efficiency tests to municipal government, but the conviction that only through efficient government could progressive social welfare be achieved …” (1912b, 100, 107). He was, of course, defending the bureau against just the sort of claims later leveled by Merriam, Waldo, and Stivers. Cleveland titled the third chapter of Chapters on Municipal Administration and Accounting, “Who Pays for Graft and Inefficient Government?” He writes, “It is commonly assumed that the taxpayer pays for graft and inefficient government. This fallacy has been skillfully used and taught by many a self-seeking demagogue…. If government is not efficient, it is not the taxpayer as taxpayer who suffers, but each individual member of the community as the beneficiary of public funds…. The subversions of revenues … is a direct loss to the weak rather than to the strong” (Cleveland 1909, 25–26). He concludes: “What is not understood by the victim of private greed in public station is that every dollar diverted from public use is a support withdrawn from community effort to protect itself against disease and discomfort due to the conditions under which the
people of a great city must live; that while taking away the supports, the grafters would divert attention from the consequences of criminal selfishness, by making it appear that his own gains are derived from those who can afford to be victimized” (Cleveland 1909, 31).

Thus, the asserted purpose of these activities was to expand the capacity to provide public services. The evidence of actual practice is more mixed. William A. Prendergast, comptroller of New York City at the time, asserted that because of improvements made, the city had reduced taxes by $10 million, $0.12 lower per $100 valuation (Prendergast 1912, 53). J. Leggett Pultz made much of a reduced cost of production and increased revenue in public utilities (Pultz 1912). If tax savings was not the primary goal of these activities, it was, nevertheless, one of the goals. In Allen’s view, these were not contradictory objectives. Reducing waste was necessary for service expansion, because the taxpayer was reluctant to pay more, or even continue to pay the same so long as there was a perception of misuse of funds received (Allen 1906).

A realistic view depends on the context. In the 1870s, the “Tweed Ring” brought local government in New York City into considerable disrepute among taxpayers (Allen 1993). Similar problems were found throughout the United States and persisted into the early twentieth century. Taxpayers were dissatisfied and resisted yet more expense (Kahn 1997). At the same time, there was growing demand for government services and for protection of the underprivileged. No realistic program for expanded government services could ignore the dissatisfaction of taxpayers.

There was a perception that government expenditure was dominated by waste and theft. Welton describes government laborers as working at a “snail like pace” and that wasted time accounted for 40 percent to 70 percent of work time (1912, 103), which Cleveland calls “theft of time and service” (1909, 26). Others believed that water inspectors spent their time in the local taverns, that “city property has been subject on occasion to surreptitious abstractions” (Taussig 1912, 60–61), and, in general, that government was getting far less for its expenditures than full value. In this context, it was not unreasonable to believe that service capacity could be expanded and that taxes could be reduced. If the latter did not occur, at minimum the taxpayer was entitled to see where his money went. Later, this view gave over to more overtly tax saving initiatives (Kahn 1997), but during this period there was a genuine belief that government could expand services and cut taxes, that is, perform more with less.

Accountability

The role of performance measurement was to reconcile administrative discretion with accountability. Unelected administrators were to be given sufficient discretion to accomplish their assigned goals, but in return, they would be monitored by elected representatives of the public, the conduit of discretion, to verify that the will of the legislature or the elected executive was being followed (Goodnow 1912; Prendergast 1912; Welton 1912). The larger goal was to implement executive management with legislative oversight, following the model of a board of directors in a business corporation (Kahn 1997; Schachter 1997). A difficulty was that the technique might not work for this purpose. Frank Goodnow, who is widely associated with the politics–administration dichotomy, which is often treated as equaling the policy–administration dichotomy, and, who is sometimes thought to be the founder of professional public administration, puts the matter this way:

Suppose now that after considering such reports and such estimates the budget making authority were not convinced that its mandates had been heeded or that the administration had been efficient, the question will naturally be asked what can the budget making authority do? There are practically only two things. These are, first, it may specify items of future appropriations in great detail with the result of so diminishing administrative discretion as to impair administrative efficiency. This is what has been done in the past. Or, second, it may express its disapproval of the administration by cutting down or refusing altogether appropriations for those services which in its opinion have been unsatisfactorily managed, and by continuing in this course until those persons in charge of administrative services who have been found wanting have severed their connection with the government. (Goodnow 1912, 75–76)

Goodnow recommends the second alternative, essentially punishing the executive branch of government by withholding funds. Yet, he sees the difficulty that this approach enmeshes the legislative branch quite deeply in policy execution. He fails, however, to identify another shortcoming. Withholding funds may serve to require more efficient administration or stop the pursuit of objectives the legislature opposes, provided those objectives have not already been achieved. However, it has no force at all when the executive’s objective is to avoid some action the legislature prefers. We can see, then, that by 1912, there were the beginnings of doubt that performance measurement could provide a method by which executive government could both have discretion to get the job done and be held accountable to the will of the legislature and the citizenry.

From Then to Now

Do the bureau’s practices form one strand of the beginning of performance measurement, or are they isolated activities that are unrelated to modern practices? Modern discussion of performance measurement can be linked to
the work of the Urban Institute since the 1970s and to the publications of the International City/County Management Association (ICMA). Bouckaert traces these publications backward through the ICMA to the publication of Ridley and Simon’s *Measuring Municipal Activities* (1938) and Ridley’s earlier work *Measuring Municipal Government* (1927), which was jointly published by the Municipal Administration Service and the School of Citizenship and Public Affairs, Syracuse University. The development before that period is not well-documented, including only one earlier performance-measurement citation, to Jesse Burks’s 1912 article “Efficiency Standards in Municipal Management.” In 1912, Burks was the director of the Philadelphia Bureau of Municipal Research, one of the earliest replications of the New York Bureau of Municipal Research. The link, however, is much stronger than this. In particular, Russell Forbes cites three major contributors to performance criteria in the foreward to Ridley’s 1927 book: Charles A. Beard, Lent D. Upson, and William Bennett Munro. Beard was the director of the bureau’s training school during the second decade of the twentieth century. Upson was the director of the Detroit Bureau of Governmental Research during the 1920s. Only Munro, an established political scientist of the era, was not directly a member of the municipal research community. Beard’s and Upson’s works are reasonably characterized as reflecting the bureau’s application of the social survey methods to the administration of government (Beard 1923; Upson 1924). Munro’s text reflects a broader approach to government performance (Munro 1926). However, even he does not reflect an independent development; Munro was the regular municipal affairs contributor to the *American Political Science Review* from 1908 through 1913 and frequently reported on the bureau’s activities. Overall, 21 of Ridley’s 95 bibliographical entries are authored by directors or employees of municipal research bureaus, and many more are indirectly linked to their practices.

By 1927, the policy-oriented issues often addressed in social surveys had become separated from the performance of government agencies addressed by Ridley. His text anticipates governmentwide performance reports, such as New York City’s annual *Mayors Management Report*, and, in fact, he and Simon developed criteria for this approach in the 1930s and 1940s. These early works also anticipated comparative community rating systems (Ogburn 1917) and citizen satisfaction studies (Bruere 1912b), although not in the form of surveys. Benchmarking was anticipated in the form of setting efficiency goals associated with quality services, not necessarily the least cost (Bachman 1912).

The bureau had a profound impact on modern administration, for four reasons. First, it was intimately connected to the emergent academic establishment. Cleveland and Allen held doctorates, and Cleveland taught at New York University. The bureau was closely linked to professors at Columbia such as Frank Goodnow, who was one of the preeminent political scientists of his era. Other significant academics who were associated with the bureau included Charles Beard, Luther Gulick, William Mosher, and Charles Merriam. Bureau writings were frequently published in academic journals.

Second, the bureau intentionally exported its work through service to other communities, the spread of written material, partly supported by the Metz Fund, and the promotion of bureaus of municipal research or governmental efficiency in other localities. The first replication was in Philadelphia in 1908. By 1912, there were bureaus in Cincinnati, Milwaukee, and Hoboken, among others. By 1916 the bureaus had spread to 15 cities, with many directed by former employees or students at the New York Bureau of Municipal Research (BMR 1916). Norman Gill (1944) charts a continuing growth up through the middle of World War II, although he gives no firm number. He also cites fiscal retrenchment during the Depression and mentions a new Cincinnati bureau in the 1920s, which suggests that some bureaus had been abandoned and reestablished. This spread of bureaus and published materials reflects a deliberate diffusion of the bureau’s agenda.

Third, the bureau promoted its ideas with government officers and administrators directly through contact of government agencies such as the New York City Health Department or through officials such as Comptrollers Metz and Prendergast, and indirectly through training of future administrators at its training school, which later was transferred to the School of Citizenship at Syracuse University. Charles Beard was the original supervisor of this school. Luther Gulick was one of its students. The bureau was also linked to the National Municipal League and to other good government groups both in New York and elsewhere.

Fourth, the bureau promoted its ideas through its role with the federal government. This began with Frederick Cleveland’s tenure as chairman of President Taft’s Commission on Economy and Efficiency. Cleveland was also directly involved in the establishment of the Institution of Governmental Research (IGR), which later became the applied branch of the Brookings Institute. Lewis Merriam, a former bureau employee, was noted for some of his studies for the IGR. During the 1920s, the bureau established and, in fact, became the Institute of Public Administration (IPA), which focused more on national and international governments. The IPA still exists, having become a component of New York University in 2000.

**Lessons For Today**

We can learn four lessons from the bureau’s measurement initiative:

---

652 Public Administration Review • November/December 2003, Vol. 63, No. 6
1. Measurement should be pursued for a practical purpose, not as an end in itself.
2. An effective measurement system requires attention to both efficiency and effectiveness.
3. Measurement of efficiency requires attention not only to outputs or outcomes, but also to inputs.
4. Measurement arises in a political and social context, which both enables its occurrence and limits how it can develop. Proponents of performance measurement must build the political support necessary for long-term viability. To avoid misuse of performance reports, users must be aware of the political constraints that determine what and how information is reported.

**Practical Purpose of Measurement**

The bureau did not measure government as an end in itself. It helped government to observe itself so that government could (1) be held accountable for its use of resources to meet public purposes (reporting); (2) allocate resources for public purposes (budgeting); and (3) get better at using resources for public purposes (productivity improvement).

The advocates of modern performance measurement are similarly focused on practical uses. According to the Government Accounting Standards Board’s Web site, “Performance information is needed for: setting goals and objectives[,] planning program activities to accomplish these goals and objectives[,] allocating resources to programs[,] monitoring and evaluating results to determine if progress is being made toward achieving the goals and objectives, and modifying program plans to enhance performance” (GASB 2001). However, this practical objective may not be adequately communicated in the current diffusion of this practice across the country, as the following example illustrates.

In 2001, an organization of professionals held a general staff meeting to discuss an upcoming outside audit. The auditors expected a report from the organization which was to contain specified elements of information. One of these elements was the organization’s performance measurements. The organization had no performance measurements, although it routinely collected data on various matters. The professionals discussed this matter for only a few minutes. Soon, a staff member announced that routinely collected data could be reported as performance measurements. No one asked what, in particular, these data were or how they might be related to the organization’s mission, which had been discussed earlier in the meeting. There was no discussion of using the measurements by management, much less external bodies. In fact, this assertion resolved the matter, and the meeting moved on to other issues.

This incident is anecdotal and may poorly represent the diffusion of performance measurement. Yet, similar incidents have been reported elsewhere in the public administration literature (Beresford 2000). The repackaging of organizational processes in the guise of new practices is a known phenomenon (Gianakis and Davis III 1998), which may reflect poor design or communication of the new practices (Williams 2000), but also may reflect a mismatch of goals between those promoting the practices and those who must implement them (Taussig 1912, 59). For performance measurement to become an effective and lasting governmental practice, its modern advocates must pay careful attention to its role in practical applications. Measurement cannot be a mysterious practice implemented because external forces have heard about it and demand that it be implemented. It must be an integrated part of organizational management.

**Productivity and Outcomes**

The second lesson involves the conceptual basis for performance measurement. The bureau developed three kinds of measurement: needs assessment, outcomes measurement, and productivity measurement. However, productivity measurement dominated. In modern literature, outcomes measurement dominates, needs assessment has moved to a different branch of study, and the occasional dominance of productivity measurement is considered a mistake.

Outcome is a political concept. While management certainly should be interested in outcomes, it is also of great interest to the elected executive, the legislature, and the public. Outcomes measurement relates to the accomplishment of social objectives. Reports of outcomes must be accessible to those not familiar with organizational processes or jargon and must relate to matters valued by the society at large.

Productivity, on the other hand, is a management concept. Productivity measurement relates the use of resources to produce direct outputs—that is, products or services—of governmental processes. Productivity measurement may involve numerous reports, the meaning of which may not be obvious to those outside the organization. Of course, political participants are also interested in productivity, but they do not exhibit long-term attention to productivity (Downs and Larkey 1986, 59–94). Ironically, external political agents often press for productivity improvement as a means of achieving tax savings. One should not confuse the objective—efficiency—with the application itself. While outcomes information may be of direct interest to members of the external environment, productivity information may hold little meaning unless mediated through interpretation.

Productivity is certainly related to outcomes, but the relationship is not straightforward. A very productive organization—that is, one that very efficiently converts resources
to outputs—may yet perform poorly if the outputs do not contribute to socially valued objectives. This enigma is at the center of the difference between the public and private sectors. In the private sector, efficient production of poorly valued outputs is unprofitable. The public sector may not experience this penalty. Further, public-sector managers may not be free to change the nature of their outputs, as they may be required by governing legislation or by the expectations of higher-ranking officials. Thus, public-sector managers may have little influence over their own performance. It should not be surprising, then, that public-sector managers may prefer to measure productivity, or even access to resources, rather than performance (Beresford 2000). This resistance is not new, as it was observed as early as 1912 (Taussig 1912, 59).

The bureau was not careful to distinguish between productivity and outcomes objectives. On the one hand, it sought to improve productivity, but on the other, it aimed to communicate accountability information to the public. To serve productivity, reports were issued frequently and contained considerable detail (Burks 1912b; Cleveland 1909; Pultz 1912). Sometimes, the bureau chose to submit its findings confidentially in order to enable an organization’s officials to make changes and take credit (BMR 1916). These management objectives were in some conflict with the public purposes. Were the performance reports to be of a character usable by the public? Or, were they to be frequent, detailed, and sometimes confidential reports for managers? The bureau saw no conflict. While they recognized a distinction between saving money and meeting needs, this distinction had not crystallized into two distinct approaches to measurement. This murkiness, as much as anything, contributed to the diversion from the bureau’s agenda of capacity expansion to an agenda of tax reduction.

The tension between productivity and outcomes has not been resolved. Attempts to integrate them into blended indexes are not readily interpretable (Bouckaert 1995, 399; Holzer 1991). When they stand side by side, they can provide confusing information, allowing every organization to appear to perform well by selectively emphasizing measures. Ignoring productivity is unrealistic in an environment of scarce resources. And ignoring outcomes would leave government programs at risk of being pointless.

Some techniques attempt to sidestep these difficulties. For example, the citizen satisfaction survey goes right to the point: Are citizens happy with what the government does? But this approach does not really measure government. Unless citizens have and use more direct measures of government, their satisfaction is no more than the vague impressions that the bureau set out to improve. A general satisfaction survey can serve a role in needs assessment, but it does little to report the performance of government itself—performance must be assessed directly.

Without measurement, these matters are too easily manipulated, which happens all too often even with measurement. Treating productivity as outcomes insulates administrators from responsibility for that which is most valued from their organizations. Meanwhile, treating outcomes as productivity can be used to bash administrators for political decisions outside their control. Thus, in the political environment, champions of results ask for more resources, while champions of taxpayers assert that organizations need to make better use of the resources they have.

The lesson is that the performance-measurement community must find a way to communicate both outcomes and productivity information so that it is understood by the ordinary user. The user should not confuse one with the other. Miserly budgets should not be mistakenly admired as productive, nor platinum programs treated as a model of government accomplishment. Public employees should not be blamed as wasteful for faithfully implementing programs that have weak political support, nor blamed for policy failure where resources are not provided. It is important to continue working on meaningful and understandable reports that address both productivity and outcomes.

Data Collection for Measurement

The third lesson from this history is found in the development of a data-collection infrastructure. The bureau built its system around the functional accounting system, the ancestor of line-item budgeting (Kahn 1997). Before ballooning into too much complexity, functional accounting allowed for the coordination of financial and performance data. While this system was a great stride forward in its time, it is now a century out of date. Modern performance-measurement literature focuses primarily on how to measure the intermediate or end products of governmental activity. The link back to resources is weak. While federal reforms of the 1990s provided some attention to cost accounting, state and local governmental cost accounting is inadequate at best (Ammons 2001b; Lee 1997). Measuring the efficient use of resources is, therefore, dependent on questionable resource data. The bureau understood that measurement must attain adequate quality in all of its components. There is little current research in governmental cost accounting. Also, in a recent survey of the National Association of Schools of Public Affairs and Administration, cost accounting ranked the lowest of 13 financial management skills taught to master’s degree seekers (Rivenbark 2001). A few years ago, there was a spurt of interest in activities-based accounting, but to be implemented well, this technique depends on an existing cost accounting system. The development of reliable cost data is a prerequisite for reporting the unit cost of governmental activities.
Politics and Measurement

Finally, measurement practices arose in a political and social environment that made them possible and set limits on what they could be. The main environmental factors in 1900 were the shift of power from the legislature and minor elective officials to the singular governmental executive in the context of tax resistance, the broad perception of extensive waste and corruption in government, and the rising status of empirical science. Through measurement, the governmental empiricist could show the legislature and the public that the singular executive was successfully eliminating waste and corruption, or expose his failure. Because of this role, measurement focused on productivity. Both internal and external observers agree on this consequence (Brueer 1912b; Cleveland 1909; Merriam 1923; Waldo 1948).

How do modern political and social conditions provide opportunities and set limits on performance measurement? Since the Congressional Budget Act of 1972, the United States has seen a weak reemergence of legislative democracy. Minor elective officials have become replaced by quasi-governmental policy entrepreneurs. Tax resistance remains strong. Government workers are seen not as corrupt, but as incompetent. And, while the empirical sciences are still admired, they are sometimes thought to be unnecessarily narrow, focused on trivia or out of control. The empiricist can no longer be the arbiter of public morality, as he or she is all too often thought to need moral instruction.

Some advocates of performance measurement treat it as inevitable and here to stay. The rise and fall of past reforms, including some that look very much like the current one, do not reassure the observer of their inexorability. Modern governmental performance measurement and productivity improvement are likely to remain significant practices exactly as long as they remain relevant to the political agenda. The practical implication is that performance measurement cannot be advanced solely on technical grounds. To remain a viable practice, it must build an institutional backing and a political constituency. Because of these political considerations, it is unlikely that performance measurement can exist independently of its political context. Thus, the drive for capacity building through outcome reporting must be balanced against the political reality of tax resistance. In the early twentieth century, this balance eventually tilted far toward efficiency, which discredited performance measurement among those who were interested in building an active government. However, this is not a reiteration of the second lesson. The point is that both the practitioner and the end user of performance measurement should attend to the political constraints that lead to the particular form of performance reports. As with its analytic cousin, cost–benefit analysis, performance measurement may provide a false sense of objectivity when political constraints are ignored.

Acknowledgments

I would like to thank Hindy Lauer Schachter, Jeanette Ellis, James Guyot, Yin Yan Wang, Romuald Litwin, Allan Campbell, Aynur Saribay, many attendees at conferences of the American Society for Public Administration and the American Political Science Association, and the editors and reviewers of this journal, for their contributions to this project.

Notes

1. At a meeting at the New School in New York.
2. For this article, “inputs” refers to the resources used by government, commonly money, but also factors that can be purchased with money such as labor, time, and, sometimes, even the population in need of service. “Output” refers to the immediate material effect of government processes. As seemingly distinct as input, output, and the intervening process may be, they are not always easily separated. A professor’s lecture, for example, may be thought of as the input of time and labor, the process of actually lecturing, or the output of information communicated. “Outcomes” are the social changes following outputs or processes—for example, the employment status of college graduates some years following a lecture. Sometimes, outcomes are distinguished from “results” (sometimes called “impacts”), the specific outcomes that are causally linked to outputs. Outcomes can be broader, including changes in society that would arise pretty much the same in the absence of some particular outputs of interest, or equally likely with many likely alternative outputs. Some people fail to make this distinction, treating all outcomes as results or restricting the meaning of “outcome” to result. To get to results, or even outcomes, one may need to examine “social indicators,” measurements of the condition of society itself, or, at minimum, conduct an “evaluation study” to examine the relationship between processes and outputs to subsequent results and outcomes. “ Benchmarks” are goals set for future outputs, outcomes, results, social indicators, or “efficiency ratios” that measure the relationship between desired ends and the resources needed to obtain them.
3. There are three reasons for this restriction: First, the bureau developed its performance-measurement ideas in the early years. While in its second decade it remained principally faithful to these ideas, it branched out into many additional activities which may have shifted its focus. Second, the bureau underwent a major transition soon after 1912. In 1910, Frederick Cleveland, one of its directors, became more closely associated with the scientific-management movement (Haber 1964). In 1911, he left the bureau for a period of time to head President Taft’s Commission on Economy
and Efficiency. Two of his fellow commissioners were Frank Goodnow and William F. Willoughby ("President's Commission" 1911). Willoughby was involved in the other major program in performance measurement during the first decade of the 1900s, that of the Census Bureau, and his subsequent tenure at the Institute of Government Research suggests a narrower conception than that sometimes exhibited by the bureau (Kahn 1997). In summary, during this period Cleveland had considerable opportunity to change his views. About the time that Cleveland returned, Henry Bruere and William Allen left (Kahn 1997). Consequently, by the middle of the second decade of the twentieth century, the bureau was not necessarily the same institution in the period under discussion here. Third, in 1912 the Annals of the American Academy of Political and Social Science published an edition devoted to the efficiency efforts of the bureau and related organizations. This publication marked the ascendency of the first period of development.

4. The following is an overview of the sources of the performance- and productivity-measurement activities of the New York Bureau of Municipal Research, with citations and explanation omitted: (1) The social survey developed by Charles Booth in England during the 1880s and imported into the United States by the settlement houses; (2) the collection of municipal statistics, primarily in Germany, and imported into the United States by the Census Bureau and its predecessor at the Department of Labor beginning in 1898, following similar developments in Massachusetts in the earlier 1890s; (3) the development of policy studies, possibly in England in the seventeenth century, with intermittent policy reports at such organizations as the Statistical Society of London; (4) the sophistication of governmental accounting, beginning around 1900, partly through the efforts of the accounting firm Haskins and Sells, where Frederick Cleveland worked for a period; (5) the increased sophistication of cost accounting during the end of the nineteenth century; (6) the common practice of annual reporting; (7) the development of scientific management, primarily after 1910; (8) the development of budgeting as a management device by Frederick Clow in the 1890s; (9) the development of statistics as a science of method, rather than its earlier incarnation as numerical sociology, throughout the nineteenth century, but primarily after 1850; (10) the European practice of municipal exhibitions; (11) the practice of rating employee efficiency which arose in the United States government in the 1850s; (12) the political conditions that are discussed in the body of this article. To all of this, one might add a general attitude of empiricism that prevailed during this period.

5. Each issue of the Annals is devoted to a single topic; this issue is devoted to efficiency in municipal government, primarily the work of the New York Bureau of Municipal Research.

6. A more extensive discussion of the Budget Exhibit can be found in the works of Dahlberg (1966) and Kahn (1997).

7. Jane Dahlberg says that, at a later period, the functional budget was discarded as too restrictive (1966, 174–99), but by then it had already evolved into the line-item budget (Kahn 1997, 88 ff.).

8. In 1910 and 1911, Louis D. Brandeis used the vehicle of scientific management to challenge the rate-setting objectives of the railroad establishment, simultaneously popularizing himself and the views of Frederick Taylor (Haber 1964).

9. In later work, Taylor somewhat distances himself from pure time and motion study (Schachter 1989), but any reasonable reading of Shop Management, Allen’s source of scientific management knowledge in 1903, places it at the center of scientific management. For example, Taylor says, “In no case, however, should an attempt be made to apply these principles unless accurate and thorough time study has previously been made of every item entering into the day’s task” (Taylor 1947).

10. In general, the impression given in this literature is that the legislature is the voice of the people and the executive, including the elected executive, is monitored through the devices developed by the bureau. See, for example, Goodnow’s article cited at this point in the text. It is not, however, clear where the bureau stood with respect to the struggles between the executive and the legislature. Allen was a proponent of citizen power and could be thought to lean toward legislative democracy (Allen 1917). Cleveland, particularly following his stint with the Commission on Economy and Efficiency, may have been more of a proponent of the executive. Bruere’s New City Government (1912b) seeks to show the flaw in the commission model of government, so he implicitly supports a strong executive.


References


