The Role of the ICMA in Promoting Performance Measurement through ‘Standards’: 1927-1956

Dan Williams
CUNY Bernard M Baruch College

Recommended Citation
https://academicworks.cuny.edu/bb_pubs/230

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/bb_pubs
Part of the Policy History, Theory, and Methods Commons, and the Public Administration Commons

This Presentation is brought to you for free and open access by the Baruch College at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
The Role of the ICMA in Promoting Performance Measurement through ‘Standards’: 1927-1956
Daniel W. Williams, Ph. D.
Baruch College, School of Public Affairs
Daniel_Williams@baruch.cuny.edu
March 20, 2004
212-802-5983

Conference Paper
Not for Publication or Citation

This project is supported by PSC-CUNY grant #656170034

Copyright © Daniel W. Williams 2004
The Role of the ICMA in Promoting Performance Measurement through ‘Standards’: 1927-1956

Abstract

This paper examines the role of the ICMA in promoting performance measurement during the years of Clarence Ridley’s administration. Clarence Ridley received his Ph.D. at Syracuse University in 1927 and soon, thereafter, became the executive director of the ICMA. His Ph.D. topic constitutes the first book specifically focused on performance measurement. It differs from other performance measurement literature of that era in that it is more instrumental and oriented towards management objectives and standards. In contrast, earlier literature, such as that of Charles Beard or William Munro, provided significant consideration to performance of government as a political entity. This paper examines the changing concept of governmental performance under Clarence Ridley’s watch.
The Role of the ICMA in Promoting Performance Measurement through ‘Standards’: 1927-1956

A principle thesis of New Public Management is that political leaders can obtain more public service out of government if they focus their communication on outcomes rather than activities. In brief, decision makers should tell public managers what they want specified in outcomes and hold the managers accountable for getting there. Holding public managers accountable for achieving outcomes is the business of performance measurement, which has been one of the most widely touted rational techniques of the last decade. Performance measurement was, of course, not invented 10 or 15 years ago, but, as I have shown elsewhere, can be closely associated with the “government research” of the good government associations of the Progressive Era (Williams 2002, , 2003, , 2004).

In this paper, I examine a development that begins or accelerates in the depression years, the development to focus on performance through “standards.” This practice is with us today. For example, Rossi, Lipsey and Freeman (Rossi, Lipsey and Freeman 2004, 75) recommend 12 different sources of “criteria” for evaluating programs in one of the better known program evaluation textbooks of today. They argue that a program evaluation cannot be completed without reference to such standards. Where do we get this practice and what is it about?
One might assume that the use of “criteria” or “standards” is an intuitively obvious and absolutely necessary component of any practice that might be called “evaluation” or that addresses the achievement of “outcomes” as it is presently called in the popular terminology. However, this direction has actual historical sources, so it is worthwhile knowing these developments. Furthermore, the term “standards” and such alternatives as “criteria” are so vague and used so equivocally as to suggest that they can sneak the very antithesis of science, raw judgment, into seemingly scientific practices such as performance measurement and program evaluation.

**Antecedent and Concurrent Conditions**

In the 1920s, prototypical performance measurement practices included: (1) citizens’ – that is, conducted by citizens or their proxies – audits (called surveys) of governments or communities to establish how the community fared in comparison to other communities (Aronovici 1916, ; Mark 1916, ; Taylor 1919); (2) a somewhat more professionalized version of similar activities conducted by government research organizations, sometimes focusing on specialized features of governmental management (Beard 1923, ; Gulick 1928, ; Upson and City Survey Committee 1924); (3) the rise of comparative community or governmental rating activities (Bracy 1924, ; Cramer 1929, , 1934, ; Ogburn 1917, ; Walker 1930); (4) the beginning of a literature in governmental performance from political science (Head 1927, ; Munro 1926); and (5) the first dissertation, quickly to become a short book, on the performance of governmental services (Ridley 1927a, , 1927b).
By the mid-1920s, many of these practices involved normative leanings. In fact, in general these practices originated with normative purposes. For example, many of the citizens’ surveys were conducted to determine community shortcomings in order to fix them. Often, these shortcomings were presumed to fall in the area of failing to meet the needs of the down trodden. Community surveys were the descendent of work from an even more intense form of community audit, also called a survey, that originated in England and was imported into the United States through the early settlement houses, and the progressive social objectives of these early activities was undisguised (Aronovici 1916, Bartlett 1928, Bulmer, Bales and Sklar 1991).

One of the developments of this period was the intense interest in comparison with other like communities. In an early example of this, William F. Ogburn had his students at Reed College develop a “survey” of communities across the United States to determine relative community standing on 17 characteristics (Ogburn 1917). This survey, as with many others in this period, did not involve collecting primary data. Instead, it used secondary data to rank communities on a variety of socially desirable dimensions from wages to church attendance. Factors that relate to modern concepts of social indicators included infant mortality and illiteracy, while matters that might be more directly associated with government production of services included matters of parks, pavement, fire loss, public properties, circulation of library books, school attendance, school property, and teacher salaries. These comparative characteristics reflected the norms of 1917.
Ogburn was pioneering a then new development, what is now called “empirical norms.” Sometimes we may be tempted to think of empirical norms as somehow distinct, antiseptic, and somehow not “normative” in the same way as moral norms. However, this early example calls that epistemological assumption into question. Although Ogburn sought to do no more than find out how 36 cities compared with each other (the standard business of empirical norms) within 17 dimensions, being ranked poorly within a dimension left a risk of opprobrium.

Other pioneers in this practice include William Bracy and Edison Cramer who conducted three serial studies in Colorado in 1923 (Bracy), 1928, and 1933 (Cramer) comparing city governments on a variety of dimensions (Bracy 1924, ; Cramer 1929, 1934). These studies differ significantly from the Ogburn study in that they may have collected some primary data and in that they focused much more directly on government, rather than on the community as a whole. While other studies are known to have occurred, only fragments exist today. There is some indication that Bracy initiated the first Colorado study replicating or partly replicating a Nevada study; however, evidence other than mention of this study has not been established.

Another way in which the Bracy and Cramer studies differ significantly from the Ogburn study is in the refocusing of standards. First of all, the great majority of categories are not only governmental, but government service. Thus, while Ogburn is concerned about the whole scope of the community, Bracy and Cramer narrow their focus to the scope of governmental units. I will have more to say about this later. Second, for governmental
services, there are two measurements: one concerns service delivery and the other concerns service cost. Thus, for example, they discuss library circulation and the cost of delivering library circulation. While Bracy and Cramer differ from Ogburn in focusing on governmental units, governmental service and a pecuniary measure of governmental efficiency, they embrace Ogburn’s approach to empirical norms. They rank cities comparative to each other and let reader infer the implications of relative adequacy. However, by making nearly half of their entire set of rankings in terms of dollars, they radically change the center of the norms, moving away from questions of whether the community has attained the quality it would like to how well it spends its resources to reach such attainment. Thus, the norms implied in the Colorado studies are much more oriented to governmental efficiency, as opposed to what might be thought of as the effectiveness standards from Ogburn’s era.

Neither efficiency nor effectiveness is strictly comparative to the bogeyman of “Scientific Management” that is often raised at about this point in the discussion of this period of public administration. If one looks back to the mid-1890s and Frederick Taylor’s work, what he discusses is two different ways to establish expectations of work accomplishment in factories. One, which he recommends, is based on an elaborate system of systematic experimentation, measurement and selection. Thus, the employee is fit to the job and specifications are set for the job based on experiments that show exactly what can be expected. The other, which he considers to be substantially defective is “benchmarking,” comparing your current work to what you did last time period or to the industry leader and trying to do a little better (Taylor 1895, ; Taylor 1947 (1903)). It does not take a lot
of mental gymnastics to see that “empirical norms” bears a lot of resemblance to “benchmarking.” So, strictly speaking, both the efficiency and the effectiveness oriented approaches of the 1920s rely on a technique that Taylor labeled unscientific because of the way in which it determined norms.

The Taylorite approach is more closely reflected in the work of Clarence Ridley. Ridley sets out to determine the best measures of government activity in four activities, fire protection, health protection, crime protection, and public works. This emphasis on technical determination is distinctive in Ridley’s approach and will be the main subject of the second section of this paper. However, before progressing to that section, it will be useful to be familiar with some other developments arising in the or around the 1920s.

An important one of these other developments is the work of William B. Munro. Munro, a Harvard political scientist, recommended 25 criteria for evaluating governments (Munro 1926). By focusing on government, he reflects the same narrowing as Ridley, Bracy and Cramer compared with Ogburn. However, unlike Ridley, and, to a degree, unlike Bracy and Cramer, his criteria are largely focused on governance, rather than on government service. Thus, he suggests criteria concerning city home rule, the city charter, popular control of government, the number of elective officials, the terms of office, the presence of a merit system, the adequacy of the revenue system, etc. Bracy and Cramer also consider governance, but to a much lesser degree (they consider the electoral participation). Louis Head implemented the Munro criteria in a study Dallas (Head 1927).
What is apparent with Munro is that the dominant form of evaluative comment on government was not, in 1926, clearly bound towards the modern practice of evaluating government service. Had Munro’s route dominated, the focus might be at the next superior hierarchal rank and evaluation might be more oriented to how well elective representatives fulfill their responsibility to the electorate or to the citizenry at large, rather than how effective organizations are in performing assigned tasks. Sometime shortly after 1927, the current normative views underlying the modern approach to performance measurement came to dominate public administration practice.

Before turning to those developments, however, there two other series of developments worth mentioning. One of these developments is associated with Mabel Walker. In 1929 she published “Rating Cities According to the Services Which Their Citizens Are Getting,” which was slightly updated the next year as a chapter in Municipal Expenditure, both of which were based on her dissertation (Walker 1929, , 1930). This study reflects an alignment intermediate between Bracy and Cramer on the one side and Ridley (to be discussed in the next section) on the other. As with Bracy and Cramer, Walker is focused on large-scale comparative analysis. That is, she indexes 160 cities across the country. Indexing is similar to ranking, but has an interval scale and in her usage it is normalized to have a maximum value of 100. Walker is also like Bracy and Cramer in using secondary data for many of her categories. However, her focus is more on effectiveness, which she calls “results,” than efficiency, which brings her closer to Ridley. Her resemblance to Ridley also arises from her lack of interest in governance.
She addresses matters of government service, only. The difficulty with Walker’s work, as with Munro and Head is that they appear to be antecedents to nothing. In 1939 and 1940 Edward L. Thorndike completed two studies (Thorndike 1939, , 1940) that appear to combine some of the precedence of the Ogburn line of thought with the Walker line of thought, but Thorndike is much more the sociologist following the Ogburn train of thought. He is looking into communities, not into government except as it is incidental to communities.

While Walker’s work appears to be a dead end in the evolution of performance measurement, it does reflect the scientization of the performance standards. As with Bracy and Cramer, Walker set standards through empirical norms. But, her standards move much more in the direction of technical concerns, for example the number of times a week that garbage is collected, or percent of felonies cleared by arrests (Walker 1930, 63). This is understandable in that her paper appears two years after Ridley’s book, which she cites (Walker 1930, 58).

The last antecedent that I will discuss is the work of Lent Upson. Based on the evidence, Upson is Ridley’s mentor. He provided documents that Ridley relied on in completing his dissertation. Ridely’s chapter on crime protection is highly influenced by Upson’s work for the International Chiefs of Police, and shortly after becoming the executive director at the International City Managers’ Association, Ridley took on an associate, Donald Stone, who had worked for Upson in the work for the International Chiefs of Police.
Unlike Ogburn, Bracy, Cramer or Walker, Upson sought to establish technical norms rather than empirical norms for work performance. In practice, technical norms are empirically determined, but the empirical approach is quite different. As I have described them, empirical norms are mere rankings. The “norm” is a rank at or near the top. In Walker’s case, exceptional values, as judged by Walker, were excluded from the rankings, but the remaining use of straight ranking remained. Upson is not so clear on this matter. He is looking for “standards,” a measure of what can normally be expected. An Upsonian standard may fall at a more micro-level than the whole organization. Upson advocates the measurement of individual workers and individual tasks. Upson’s work also reflects more of a management perspective, rather than a citizen perspective. As a manager, he is looking at whether the work unit is accomplishing its role.

Upson unabashedly advocated the perspective of scientific management. From this point of view, he wanted to know how individual workers contribute to the accomplishment of the work units’ goal. Implicitly, if an individual worker is not contributing adequately, corrective action is required. Thus, work should be measured and work below the standard should be corrected. The standards should reflect technical feasibility in the usual case, not necessarily the highest possible level of accomplishment as is often the target with benchmarking approaches.

In a study of Cincinnati and Hamilton County, Ohio (Upson and City Survey Committee 1924), which Upson led, the study committee attempted to measure work from every
functional activity of government. Many of these measures were weak or ephemeral. Nevertheless, the message that was implied was to accomplish a high level of work output for resources consumed. Yet, in one way this study reflects the pre-Ridley perspective. That is, a small section of this study looks at citizen access to government, a governance issue.

**Clarence Ridley in 1927**

In 1927, Clarence Ridley published *Measuring Municipal Activities* (Ridley 1927a, , 1927b), which remains, today, the intellectual foundation of modern performance measurement in the United States and the world. Ridley is the intellectual heir to Upson. While his text is brief, his approach is technical. First, he examines the technical literature of performance measurement from fire protection and health. In 1927, these are the two main areas of public service that have a technical literature of measurement. Next, he provides a pre-release version of the work of the International Chiefs of Police on crime reporting. His close association with members of the International Chiefs of Police task force, particularly Upson, enables his work in this area. Finally, he provides the first cut at some work he further develops from 1928 through the early 1930s concerning measurement of public works activities. What is unique about Ridley’s work is its technical focus. While the book is brief at 80 pages plus a brief bibliography, it discusses issues of measurement concept, addresses only four types of measurement, attempts to settle matters between competing measurement approaches, and makes no attempt to actually implement the measurements. It is, thus, the intellectual foundation of performance measurement practice.
Ridley’s work is parallel to Munro’s work, in that Munro also examines concepts rather than trying to implement a practice. However, while Munro develops 25 criteria for government practice, which he partly operationalizes preparing the way for Louis Head to try to actually observe a government on these dimensions, Ridley is focused on the technical success of specific measurements of specific services.

Why does this matter? Ridley’s technical focus has an entirely different normative framework from that of any of the predecessors mentioned, with the possible exception of Lent Upson. Ridley’s approach is the science of management, or the science of administration. He aspires to absolute norms, for example in 1943 he and Herbert Simon say that efficiency refers to the “ratio of the effects actually obtained with the available resources to the maximum effects possible with the available resources (Ridley and Simon 1943b, 3 (Italics in the original)).” The phrase “maximum effects possible” is related to an absolute, not the sort of empirical norms that satisfied everyone from Ogburn to Walker. That does not mean that Ridley would not use empirical techniques to observe the maximum possible effect, but the tone of this phrase reflects Taylor’s view that there is an establishable technical level to which the work unit or the worker can be held accountable.

The effort to establish technical norms brings us to the word in the title of this paper, “standards.” The role of the word “standards” is not that it differs from other norm setting concepts in setting norms, it is that it implies a difference in the source of those
norms. In particular, standards are the kinds of norms that are set for technical reasons.

In an industrial setting, the purpose of a process may be to attach a desired characteristic to an object. A “standard” can be set regarding, how well attached the characteristic is, how efficiently the characteristic is attached, or regarding the quality of the characteristic once attached. For example, a ball bearing might be given the characteristic roundness, but if the ball bearing is made of soft metal, the characteristic may not be very well attached, since slight pressure might make the ball bearing flat. Alternatively, roundness might be very well attached but at an extraordinary cost. Finally, one might find that the ball bearing is reliably “round” at a reasonable cost, but not really all that round.

Standards can be set for any of these. From Taylor’s perspective, using Ridley’s terminology, there is a maximum possible.

In the public sector, these “standards” are normatively problematic because they seem to be neutral. That is, the supposed real norm comes at the beginning, when we decide what the program is intended to do. However, there are other hidden normative implications that get involved. I will discuss one involving a Ridley editorial in the next section. Through the standards’ implied requirement for accomplishment at any cost, they suggest conflict with, if not overriding of, any contrary moral judgments. When public decision-making is faced with other sources of values, such as equity or public preferences, these standards can become barriers.

Few familiar with public administration literature will be surprised that there is a conflict between values that arise from public preferences or values implied when public goods
such as jobs must be distributed equitably and values implied in efficient and effective delivery of particular services. However, the typical discussion leaves the locus and rationale for these conflicts murky. What is apparent here, is that the management rationale represented by the growing strength of the Upson-Ridley branch of thought is directly related to this conflict. This branch of thought was one of several competitors during the 1910s and 1920s. What changed?

The International City Managers’ Association

In 1928, Clarence Ridley was selected as the executive director of the International City Managers’ Association (ICMA). The ICMA was formed in 1915 only a few years after the origin of the city manager profession in the United States. Ridley had been a city manager in West Virginia in the early 1920s before completing his Ph.D. at Syracuse University in 1927. His appointment as executive director of ICMA was concurrent with an academic appointment at University of Chicago and with the relocation of the ICMA’s headquarters from Lawrence, Kansas to Chicago. His core assignment was to upgrade the technical quality of the support flowing between city managers and flowing for the ICMA to the city managers. For example, he added academic content to the city managers’ main trade journal, which addition was an explicit expectation at the time he took on this role (Vogel 1967).

Shortly after his appointment at ICMA, we find Ridley associated with the National Committee on Municipal Standards (Stewart 1950). This committee has an interesting history of its own, which is beyond the scope of this paper. However, its essential
characteristic for this paper is that it became the vehicle for Ridley to carry out his program of setting standards for municipal street cleaning and sanitation. It also associated Ridley with Donald C. Stone, the, by then, former assistant director of the Committee on Uniform Crime Records, another protégé of Lent Upson, and with Arthur Eugene Buck, a protégé and colleague of Luther Gulick. Buck later became the chief author of the first Hoover Commission’s report on budgeting, in which performance budgeting was invented, at a time Donald Stone when was a senior member of the Bureau of the Budget.

But all of that is later. In 1928 through 1930, they developed a set of recommendations on standardizing the measurement of various forms of public works activities for performance measurement, management, and cost accounting purposes. Here, the social goal, is taken as given, e.g., that streets are to be cleaned or that garbage is to be collected, etc. Measurement is in terms of how well this objective is met through public resources (Ridley 1929, ; Stone, Moe and Ridley 1930, ; Stone, et al. 1931a, , 1931b). These texts embed their normative component so deeply that the reader may be left with the impression that they are purely scientific or technical. Nevertheless, the organization that initiated these studies was looking for municipal standards and a typical reference to these documents refers to measurement standards (Stewart 1950, 150). The standards found here require a bit of unpacking. First, there is a goal that measurement is standardized, that is made consistent across organizations, so that there can be comparison. This sort of standard, or rule, is used to make it possible to set another sort of standard, an expectation of what counts as doing a good job. Until we know which
organizations are doing better than others, we have no way of knowing which might be model for empirical investigation. In the realm of public works in 1930, matters were so rudimentary that it was not possible to do any comparative analysis, because there were no consistent units of measure. The “standards” Ridley wanted were to make this possible. Only once units of measure were consistently used could any one conduct a scientific comparative study. As Ridley eventually turned towards other matters, his normative view with respect to public works practice is not revealed.

But, Ridley did keep working on related performance measurement issues. In fact, it is largely due to him that the practice survived the 1930s and 1940s. As the director of the ICMA and the editor of its chief journal, *Public Management*, as well as the Reports Editor of the *National Municipal Review*, Ridley was in an opportune location to influence the development of public administration practice. This opportunity would be apparent even in today’s world, but it is much more significant in the 1930s and 1940s. First, these were important expansion years for the ICMA. In 1930 there were likely 500 cities in the United States and Canada with city manager plans, by 1960 this number had increased almost five fold (Vogel 1967). Second, and more importantly, before the New Deal was in full swing, local government was where public administration occurred. Due to Constitutional restrictions, the federal government still conducts much of its domestic policy through proxy use of state and local government and states leave many activities to localities. The central role of the ICMA put Ridley at the center of governmental practice. In his early and middle years, he relied heavily on his own academic interest in fulfilling the goal of upgrading the academic and professional skills of city managers.
In 1932 and 1933, these skills were put to a different use, reducing the cost of city government. In 1932, Ridley and his assistant director, Orin Nolting, made a checklist of factors that cities might consider for cost cutting. By 1933, this checklist was updated and converted into a book on cost cutting (Ridley and Nolting 1933, ; Ridley and Nolting 1932). These Depression issues would naturally dominate the activities of the ICMA in the early 1930s. At present the evidence is not in hand to evaluate, but they may have reinforced the turn towards management goals of efficiency and effectiveness. Still, these cost reduction goals are perfectly consistent with the efficiency objectives underlying Ridley’s own focus in public works. While *Measuring Municipal Government* shows concern for effectiveness with fire protection and health protection, and at least an awareness of the concept of effectiveness with police protection, his discussion of public works is almost entirely focused on volume of work produced. The same is true of the subsequent work out of the National Committee on Municipal Standards and its descendents. The practice measurement moved strongly in the direction of efficiency, period.

**Measuring Municipal Activity**

In 1937 and 1938, *Public Management* published an update to Ridley’s 1927 *Measuring Municipal Government* in serial form, with each article authored by Ridley and his research assistant, Herbert A. Simon (Ridley and Simon 1937a, , 1937b, , 1937c, , 1937d, , 1937e, , 1937f, , 1937g, , 1937h, , 1937i, , 1937j, , 1938b, , 1938c, , 1938d). In 1938,
these articles were collected along with a bibliography and a section proposing a format for municipal reports as *Measuring Municipal Activity* (Ridley and Simon 1938a), which was reissued with a new introduction and without the format for reports in 1943 (Ridley and Simon 1943a). The format for municipal reports was issued as a separate pamphlet for monthly reports in 1943 (Nolting 1943) and as a separate booklet for annual reports in 1939 and 1948 (Ridley and Simon 1939, 1948).

The first two chapters of *Measuring Municipal Activities* concern general issues of measurement and reflect more sophisticated thought on the general issues raised of the first two chapters, but particularly chapter two, of Ridley’s *Measuring Municipal Government*. The next four chapters cover the same four service categories that Ridley previously discusses: fire, health, crime, and public works. Here *Measuring Municipal Activities* is somewhat less detailed and less sophisticated than *Measuring Municipal Government*; however, it is more up-to-date, reflecting the developments of the intervening decade. These developments include the work of International Chiefs of Police and the Federal Bureau of Investigation in uniform crime reporting, the work on public works discussed above, work of national membership bodies related to public health concerning health services delivery standards, and updated standards from the National Board of Fire Underwriters. Thus, the first six chapters of *Measuring Municipal Activity* are a more up-to-date and less academic version of *Measuring Municipal Government*. 
*Measuring Municipal Activity* continues by expanding the categories of municipal services discussed to include recreation, welfare, education and libraries. These areas reflect some development, but not as much as those of the original four areas explored in *Measuring Municipal Activity*. The category of libraries is discussed somewhat further in section on standards below. *Measuring Municipal Activity* also adds a new type of measurable: staff services including personnel, finance and city planning. These are services that function to make government itself operate well rather than to meet specific public service needs. The measurement of these services remains primarily an area for future development.

The principal point is that what Ridley and Simon do in 1938 is extend Ridley’s work from 1927. The important matter for the development of performance measurement is the concepts of inputs, effort, outputs, outcomes, effectiveness and adequacy. These words are not always the ones used, but the constructs are recognizable in the words that are used. For example, the term “outcomes” never appears, but the term “results” is generally used for the same purpose. I say, generally, because this term is not always used consistently. Efficiency is used to refer to the “ratio of the effects actually obtained with the available resources to the maximum effects possible with the available resources” (Ridley and Simon 1943b, 3 (Italics in the original)).” The unwary may not realize that this is a very sophisticated double ratio, (Effect/Resources)/(Maximum Effect/Resources). Resources can be cancelled (Ridley and Simon seem to hold resources constant thereby justifying this cancellation), so the ratio becomes Effect/(Maximum Effect), but that does not change the origin of construct. Today the
more common notion of efficiency is the ratio Effect/Resources and one might think of one process as more efficient than another if the ratio is larger, but the public administrationist would not normally work with the double ratio with its implied maximum value of one.

This assumption of a maximum value is what keeps our attention on standards and benchmarking. If efficiency is simply a matter of doing better than before – which is the current notion of economic productivity used in the overall economy – standards or benchmarks would be unnecessary. One would simply need to know what she or he did last year to determine how to do better this year. However, to use a teleological approach, one must know the goal. The goal is not set as an arbitrary promise nor, in principal, is it an accident of history; it is set based on the maximum possible. While Ridley and Simon compromise and examine “relative efficiency” (p. 3) they have set out in a different direction. This direction, standards, is the heir of the scientific management. The “maximum possible” is the conceptual twin of the time and motion study. Frederick Taylor used time and motion studies and observation to determine the actual achievable work in the factory, and paid piece-rates based on achieving these “scientifically established” work standards. He specifically rejects benchmarks (prior levels of accomplishment) as unscientific and easily manipulated (Taylor 1895, ; Taylor 1947 (1903)). In the efficiency component of Measuring Municipal Activity, Ridley and Simon adopt the same conceptual frame; however, when they move to “relative efficiency” they compromise Taylor’s standard.
Ridley and Simon also raise the issue of the six kinds of efficiency:

“The term *effect* as used here includes any effort, performance, or result. The term *resource* as here used comprehends money expenditure, effort, or performance considered as productive of effect. The efficiency of accomplishment can therefore be measured in its relationship to several factors: expenditure, effort, and performance. These may be illustrated with reference to street cleaning. [1] There is an efficiency of accomplishment of results relative to expenditure (could cleaner streets have been obtained with the same expenditure?); [2] relative to effort (could cleaner streets have been obtained with the same number of man-hours and equipment-hours of effort?); and [3] relative to performance (could cleaner streets have been obtained by performing the same number of curb-miles of sweeping?). [4] There is also an efficiency of performance relative to expenditure (could more curb-miles have been swept with the same expenditure?); [5] and relative to effort (could more curb-miles have been swept with the same number of man-hours and equipment-hours?) [6]

Finally, there is efficiency of effort relative to expenditure (could more man-hours and equipment-hours of effort have been obtained at the same expenditure?) In dealing with specific problems it is necessary to specify which type of efficiency we are attempting to measure” (page 3, numbers added, Italics in original).

Ridley and Simon do not give us much to go on about all these types of efficiency except to be careful that we know which one we are measuring before we start. But, of course, they are substantially different matters. Efficiency₁ (subscript to correspond to the
numbers inserted into the quote above) is overall efficiency or how well our economic resources managed to get the sort of thing society wanted. It seems to comprise Efficiency2, Efficiency5, and Efficiency6, or Efficiency3, Efficiency4, and Efficiency5. This is a little confusing. It seems illegitimate to get, for example, more performance, Efficiency4, and more effort, Efficiency5, for the same dollars, but Ridley and Simon do not clear this up. The likely upshot is that relationship of all these efficiencies need to be worked out and the manager needs to understand which ones serve, like primary colors, as the ones that take precedence in thought. I have found no literature that provides a critical examination of these ideas.

The alert reader will have noticed another word that is no longer commonly used in the literature of performance measurement, “adequacy,” which refers to “the absolute measure of accomplishment” (page 3). Having first defined adequacy, Ridley and Simon do little more to examine it. Their message is that measurement is about efficiency. Elsewhere, however, Ridley, the editor of Public Management, makes his position clear through an editorial. Issues of adequacy are reserved for political decision makers: “[One] must still ask on what basis the specialists determine for any community how many books it wants to read, how many parks it wants to read, how many parks it wishes to have, or how low it wishes to keep its fire loss” (Ridley 1943). As the chief spokesman for the principal form of municipal executive throughout the United States, Ridley is directing the subject matter specialist such as the librarian, director of recreation or the fire chief to leave matters of level of service desired to his or her political superior, beginning, of course, with the city manager. The city manager’s role in this resource
allocating matter is left unaddressed, but it is in fact problematic. The pragmatic political theory that justified the city manager would provide for the manager to propose an allocation, which the city council could then revise. This theory initially embraced a strong mayor model of government but quickly transferred the presumed virtues of the mayor to the city manager in the mid-1910s. Thus, Ridley’s editorial amounts to telling employees one or two levels below the city manager to stop meddling in the city manager’s affairs.

**Varieties of Standards**

Ridley’s editorial hints at a developing problem with the ubiquitous idea of standards. There were so many ways to go about setting standards, but not all standards were equal. Some library standards from this period include the number of books circulated per user or citizen, volumes held per citizen, and dollars appropriated per citizen. At a different level, there was an attempt to classify reference questions by the level of difficulty of answering and an effort at classifying circulation by the inherent value of the text (more for non-fiction than for fiction). For recreation, there was a standard of play area for number of children. This area might be adjusted for type of equipment installed or age of the children. A welfare workload standard study associated number of cases per employee with number of claims disallowed, the point being that the apparent savings associated with overloading caseworkers with too many cases was penny wise, as the net effect would be incorrect allowance of costly welfare claims (Simon, et al. 1941).
It is curious which of these standards are revealing of underlying social norms after 60 years. The workload standard stands out. A seemingly technical matter, how many cases a worker can handle without making a lot of errors, masks another value determination, one still with us today, which is that we need to be careful to avoid unnecessary largess in our public assistance programs. How many more of these “technical” standards are indirectly measures for underlying social norms?

What one should note among this great variety of standards is the general lack of efficiency or effectiveness objectives of the “standards.” For example, dollars spent on the library per citizen is what would now be called an input measure, not even an output measure and clearly not an outcome measure. What it reflects is the political popularity of the library, not its accomplishment of anything. Volumes circulated may appear to reflect something approaching outcome, but it is outcome only in the private economic sense, the library as a private enterprise. For ‘volumes circulated’ to reflect a policy outcome, one would need a clearly articulated policy objective associated with this quantity, perhaps a concern over lack of access to reading material. Otherwise, it is an intermediate quantity (output measure) on the way to a policy objective (such as, literacy).

The consequence of using these kinds of “standards” is to “scientize” preferences, that is, to give a neutral scientific appearance to the policy choices of standard setters. This tendency might be a conspiracy, as perhaps Ridley appears to have thought when he wrote the editorial cited in the previous section, but it can equally arise out of confusion.
over what a standard might be. One does not need a lot of imagination to see a scenario such as…

… A table surrounded by a group of cash starved librarians discussing “standards” such as what might demonstrate how well you serve your community. The typical circulation concept is on the table, and someone asks whether circulation of *Teddy Bear of Bumpkin Hollow* (Boucher 1948) is really the same as circulation of Einstein’s *The Meaning of Relativity* (Einstein 1955). Someone else remarks, that both texts satisfy readers, so sure, but she has been wondering how to count the “circulation” of material taken off the shelf and read in the library during the day. Her library now has signs asking people not to re-shelve books, while the explanation given to the public is that librarians will be more effective in putting books back where they belong, re-shelving also provides an opportunity to count this “circulation.” Her question is, does this sort of circulation have to go into a different category than checkouts. About this time, a rural librarian begins laughing uncontrollably. Everybody stops to find out what the joke is. Finally, she says, she just cannot believe that someone would waste that much of the library’s resources on artificially producing the opportunity to count books. Her citified neighbor is probably spending more on re-shelving books than she gets to spend on buying books. The committee stops in horror. But, one by one other librarians from poor localities agree, not all from rural localities either. The chairman says, “Well, what is standard for, but to see that our patrons are well served.” After a few minutes, the first standard the new
committee sets is an expectation that the community spends at least $1.00 per

citizen on purchasing library books every year.

This scenario suggests that standards were not necessarily well defined or particularly
scientific. To cordon them into the service of efficiency would have been difficult as
they arose independently of their use by city managers. The health and fire insurance
standards were extensively developed when Clarence Ridley first discovered them.
While they were roughly consistent with his objectives, they never matched his purposes
exactly. The objective of fire loss standards was to minimize the risk of fire insurers.
This risk was almost exactly identical for building owners, so the rating system for the
fire insurers worked pretty well for the building owners and, practically speaking, for the
cities, as well. If the rating system effectively measured a decline in risk, it was
indifferent as to who bore that risk. Also, if the insurance rate declined faster than the
taxes rose to pay for improvements that reduced the risk, the citizen could see a clear net
benefit to himself despite the rising taxes.

Health standards were developed by independent professional groups, who were
primarily motivated by their interest in service of the social objectives of their profession.
Their professional status made them less susceptible to Ridley’s criticism, but in fact their
objectives, such as reducing infant mortality a typical focus of health standards, were no
different in kind from the goals Ridley’s editorial criticizes. Reducing infant mortality is
a matter of “adequacy” in Ridley’s language, not a matter of efficiency; although, with
any given amount of resources one can reframe the matter into one of efficiency.
Part of the reason that standard setting groups do not make “standards” equal “efficiency standards” is that the idea of standards had, by the mid-1920s, already had considerable development among the socially active in association with scorecards. A scorecard was a sophisticated form of a “social survey” or study of the social conditions of the local community. The “social survey” could be general or it could be focused on a particular area, such as health care. The survey developed into the scorecard, providing points for each item observed. The points required two types of judgment, one involved weighting, comparing the one item against the others in the survey. The other involved assigning actual points to a particular case. For scorecards, standards were associated with selecting items to be on the survey and with making these two kinds of judgments. The specialized health scorecard became the instrument for setting standards that Ridley latter recognized. Thus, scorecards are closely linked to standards. But, “standards” in the context of scorecards, has little to do with “standards” as Ridley wants to use the term related to efficiency.

Empirical standards/benchmarking. Standards can be broken down into the object of the standard and the value given. The object might be volumes of books circulated per citizen. The value might be 10 per year. The choice of objects appears to have been a very judgmental and, as already discussed, could be chosen with respect to either policy outcomes or to objective efficiency. The values, themselves, appear to be set in two ways, either by a priori judgment or by empirical estimate. Thus, the standard of one dollar per citizen for library services reflects the judgment of the library standard setting
group. On the other hand, certain volume-per-citizen standards appear to reflect empirical benchmarks from similar communities. The welfare workload standard is particularly interesting here. The implied objective is minimization of welfare payments, which is empirically demonstrated to be achieved through decreasing caseloads. The principal investigator for this study is Herbert Simon.

Conclusion

This paper has traced the development of standards over a 30-year period from the early 1920s until the early 1950s. During this period, the value basis for judging government changed, from more broad based social objectives in the earlier period, to narrower management objectives in the later period. The shift appears to occur fairly rapidly at the beginning of the 1930s. At this time there was a confluence of events. The ICMA was seeking to put public administration on a more professional basis and had employed Clarence Ridley for precisely this purpose. At the same time, the Depression may have rendered management objectives more critical for local government. Not examined here, there is consideration of the influence of external money, particularly the Spelman Fund, which is well documented to have played a role in the professionalization of the ICMA (Stewart 1950, ; Vogel 1967).

This shift towards “standards” was consistent with the political philosophy of the policy-administration dichotomy, which was substantially beneficial to the success of the ICMA agenda. However, because “standards” was both ubiquitous and vague, it is not clear that
hierarchical superiors gained control over decisions through the use of this construct. In effect, subordinates used the construct as a way to maintain political power. There is some evidence that this practice led to conflict between layers of municipal management.
Bibliography


Beard, Charles Austin. 1923. The Administration and Politics of Tokyo. New York; Macmillan Co.


Ridley, Clarence Eugene. 1929. Units of Measurement for Street Cleaning, Refuse Removal and Disposal With Suggested Forms for Preparing Administrative Work Programs and Reporting Actual Costs Together With a List of Factors Influencing Unit Costs in Different Cities (Tentative Draft). National Committee on Municipal Standards, Chicago.


Ridley, Clarence Eugene and Orin Frederyc Nolting. 1932. How to reduce municipal expenditures; a check list of specific suggestions for constructive economy in municipal government. Chicago; International City Managers' Association.


Ridley, Clarence Eugene and Herbert Alexander Simon. 1939. Specifications for the annual municipal report: suggestions for the content, preparation, design, publication and distribution of the annual municipal report. Chicago; The International City Managers' Association.


Upson, Lent D. and City Survey Committee. 1924. The Government of Cincinnati and Hamilton County - A report to the Republican Executive and Advisory Committee of Hamilton County. Cincinnati; City Survey Committee.

Vogel, Donald Barry. 1967. The Origin Growth and Development of the International City Manager's Association. Doctoral, University of Iowa, Iowa City, Iowa.


1 We should find Clarence Ridley’s relationship to the science of administration interesting since he later mentored Herbert Simon.

2 The public sector analogy is (1) Education that does not attach: teaching to the test. (2) Education at any cost: One-to-one tutors-to-students for every student with no limit on the amount of resources that the tutor can order. (3) Education that is not all that much education: rote learning instead of learning of thought processes necessary to solve problems.

3 The general nature of one of the normative problems is that there is a separation or alienation of decision-making from personal agency in performing public service. One does not have to treat this alienation as deliberate to see it as a problem.