

Pragmatism and Performance Measurement:
An Exploration of Pragmatic Practices in Texas State Government

by

Timothy L. Wilson

An Applied Research Project (Political Science 5397) Submitted to the
Department of Political Science
Southwest Texas State University
in partial fulfillment of the Degree of
Masters in Public Administration

(Fall 2001)

“You can’t manage what you can’t measure” has become a mantra in public administration. Over the last 10 years, performance measurement has become part of the fabric of government. But, at the same time, it has been declared both boon and bane. Which is it? It depends on whom you ask, but in final analysis it is both.

The original impulse for performance measurement carries back a century or more. At the turn of the last century, the American philosophy of pragmatism fueled the application of scientific principles to social progress, and resulted in the striking reforms of the Progressive Movement in the early 1900s (Zanetti and Carr, 2000). The field of public administration, formed as a recognizable entity about the same time, did not embrace pragmatism for reasons that will be discussed below. Recently, however, some public administration scholars have suggested that American pragmatism could, and even should, provide the field with a theoretical basis. Some scholars have gone so far as to suggest that pragmatic principles are already present in the practice of public administration, a so-called “implied pragmatism.”

One area where pragmatism and public administration practices might readily entwine is the burgeoning field of performance measurement (Shields, forthcoming). A review of the literature, in fact, indicates many commonalities between the best practices of performance measurement and the principles of pragmatism. Many of the tenets of pragmatism are paralleled in the performance measure literature. However, the literature also indicates that these “pragmatic” best practices are not always followed, resulting in considerable dissatisfaction among practitioners.

The purpose of this applied research project is to determine the extent to which these pragmatic principles are implied, indeed practiced, in the State of Texas performance measurement system.

In the following, a system of inquiry is constructed by first examining the literature surrounding American pragmatism and performance measurement. From the literature review, certain working hypotheses regarding performance measurement in Texas are distilled. Then, a method to test these hypotheses is developed and executed. Finally, results, conclusions, and recommendations are reported.

American Pragmatism

The turn of the last century marked a seismic change in American government. The proximate cause of this earthquake was, oddly, Darwin’s theory of evolution. The wide acceptance of his theory spawned an intellectual revolution as Darwinian and scientific principles were applied to many realms. One such application is the uniquely American philosophy known as pragmatism (Current, Williams, Freidel, and Brinkley, 1983, p.537-538), which flourished in America in the early 1900s (Snider, 2000a, p.332).

The philosophy that became known as pragmatism was founded over time by four primary contributors: Charles Sanders Peirce, William James, John Dewey, and Jane Addams (Shields, 1998, p.201). Garrison (2000, p.468) describes pragmatism as the union of “artistic creation with theoretical knowledge in the context of concrete practice.” Pragmatism would seem to be a blend of art and science. Shields (1998, p.201) defines pragmatism as follows:

Pragmatism is the philosophy of common sense. It uses purposeful human inquiry as a focal point. Inquiry is viewed as a continuing process which acknowledges the qualitative nature of human experience as problematic situations emerge and are recognized. Recognition involves the doubt associated with questioning existing belief systems. Doubt is resolved through critical reasoning and ultimately tested in action. It is the philosophy of common sense, because actions are assessed in light of practical consequences. Finally, inquiry is not necessarily limited to individual effort; rather it often incorporates a “community of inquirers.”

Pragmatism, then, is about challenging existing assumptions by putting them to test in the real world. It is about reconciling theory with everyday experience and practice. It is about both ends and means. Pragmatism is about identifying a problem, conducting scientific inquiry, and opening the inquiry to others in a mutually responsive fashion. It is about ever continuing this experimental process.¹ It is about treating problems in the context of the whole, mindful of the greater scheme (Evans, 2000, p.314). Pragmatism is a process; the pragmatists were a force for change.

The philosophy of pragmatism rejects the notion of absolute truth, instead defining truth as provisional. According to Peirce, truth could be found only after arduous inquiry and the passage of time (Shields, forthcoming). This provisional truth is tracked by an ongoing scientific experimental process, critical reasoning applied to the subject at hand. Problematic situations are identified. Working hypotheses are generated through the comparison of theory and practical experience (Shields, 1996, p.396-397) and then operationalized in experiment. The resulting conclusions are drawn only to be challenged again (Snider, 2000a, p.331-332). Truth is a process, not a state (Shields, 1996, p.396-397). Pragmatic truth lies in the concrete (Sahakian, 1968, p.256-259), but a peculiar concrete that can be re-formed with ease, or may just crumble into dust.

Finally, a word about the community of inquiry. Shields (forthcoming) cites three major components of a community of inquiry. The first component is the presence of a problem to be solved which draws the community together for that purpose. Second is the use of a scientific, experimental process to develop working hypotheses, and then the collection and interpretation of data to test those hypotheses. The last component is an attitude of mutual responsiveness within the community. This final point deserves further clarification.

According to Dewey (1939, p.340-344), participants should come together in an atmosphere of participative democracy. That is, the various parties, no matter their standing in the community, should be heard, their thoughts considered by the community, each party learning from the other’s viewpoints. When a member of the community has pre-formed conclusions, and will not be dissuaded by new evidence (Peirce’s notion of “fixated beliefs”), the community will not function properly (Shields, forthcoming).

Public Administration and Pragmatism

Pragmatism was in vogue around the time the first writings regarding public administration as a recognizable field began to appear (Snider, 2000a, p.335). These early writers must have thought about how their budding field should fit within the larger context of government and politics. They settled on a strict

separation between politics and administration, ends and means (Garrison, 2000, p467-468). This separation made administrators the instruments of politicians; their task was to carry out political directives in a neutral fashion. In making that separation, the field concerned itself with expertise and efficiency, not politics and theory (Snider, 2000a, p330). This separation became known in the field as the politics-administration dichotomy.

It would seem, at first glance, that public administration and pragmatism would make a companionable pairing. Indeed, many of the early public administration writers saw themselves as pragmatists (Snider, 2000a, p.336). Considering the pervasive political corruption of the time (Snider, 2000a, p.333), the notion of concentrating on the practice of public administration while ignoring the arguably grand “truths” of the politician must have been appealing. Snider (2000a) and Waldo (1984) suggest that these early writers (notably Frank Goodnow and Charles A. Beard) may have settled on some of the popular notions of pragmatism and missed the larger scheme. Their emphasis was not on experimenting but instead on developing expertise: the proper practices and principles of administration. By focusing only on practice, they missed out on one of the promises of pragmatism, the integration of practice and theory. Pragmatism did not take hold in public administration (Snider, 2000a, p.335-337).

Ironically, the dichotomous way of thinking that did not allow room for pragmatic thought in public administration has largely been discredited (Shields, 1996, p.393-394; Ellwood, 2000, p.322). Public administration became a legitimate course of university study in the 1920s and 1930s, bringing attention to the philosophy and theory of the field. In the 1940s, Herbert Simon, Dwight Waldo and other scholars, in proposing new theories about public administration, began to chip away at the dichotomy. Although not always in agreement with each other, these scholars combined to devalue the legitimacy of the administration-politics dichotomy (Snider, 2000a, p.335-349). But, no dominant paradigm emerged to take its place.

In spite of these scholarly efforts, the notion of the politics-administration dichotomy endures (Ingraham and Donahue, 2000, p.298). Ellwood (2000, p.322) cites the absence of a guiding philosophy to replace the dichotomy as the primary reason for its persistence. It is often still assumed, as a legacy of the dichotomy, that public administrators should have no voice in policy (Shields, 1996, p.394).

Several contemporary scholars have suggested that pragmatism could assume a guiding role in the practice of public administration (Shields, forthcoming; Evans, 2000). The notion that pragmatic principles have crept into the practice of public administration has also been advanced.² Performance measurement is one area where there seems to be significant implied pragmatism (Shields, forthcoming). This idea will be addressed in detail below.

Best Practices in Performance Measurement

Performance measurement can be defined as “the regular collection and reporting of information on the efficiency, quality, and effectiveness of government programs” (Martin and Kettner, 1997, p.18). Innocuous as this definition may sound, it is now impossible to hide from performance measurement in government and the public sector. Federal legislation makes performance measurement a requirement for all federal departments

(Martin and Kettner, 1997, p.25-26). Nearly all 50 states have adopted some form of performance-based budgeting (Melkers & Willoughby, 1998). Grantees and contractees receiving public money frequently have performance-based contracts (Martin and Kettner, 1997, p.21).

Performance measurement can serve at least two broad functions in government. Performance measures can be linked to budgeting, a process known as performance-based budgeting (PBB). When so linked, measures can be used at the end of the funding cycle to hold fundees accountable, or, as performance funding, to determine allocations for the upcoming fiscal year (Jordan and Hackbart, 1999, p.69).

Performance measurement data can also be used to provide program managers with information needed to adjust program activities, a process known as performance management (Heinrich, 1999, p.363). Performance data allow managers to monitor results and progress on a continuing basis, and offers staff clear direction regarding the most important objectives.

Over the last 10 years, performance measurement systems have become increasingly sophisticated. A body of literature has developed over this time which offers many insights regarding the best practices in performance measure development.

A fundamental change in performance measure development has been the decreasing reliance on inputs, the resources that go into a governmental effort. Instead, the focus has turned toward outputs, the activities or services provided. More importantly, a great emphasis is now placed on measuring outcomes, the measurable results of activities and services (Government Accounting Office [GAO], 1993, p.3). By emphasizing outputs and outcomes, measures reflect results rather than merely describing the resources committed to a given activity.

In constructing a measurement system, it is important that the combined array of measures accurately represent the total effort of a given unit. Without a comprehensive array of measures, agency staff can lose sight of the overall mission and become lost in narrow compliance (Kravchuk and Schack, 1996, p.349; Osborne and Gaebler, 1992, p.358).

According to Osborne and Gaebler (1992, p.357-358), other important components or conditions for a successful performance measurement program include broad stakeholder involvement and an iterative, evolving process. In addition, a commitment to the process of scientific inquiry (Williams, McShane, and Sechrest, 1994) and management support (Allen, 1996, p.11) are seen as requisite.

Performance Measurement and Pragmatism

A review of the discussion thus far seems to show a number of commonalities between the best practices of performance measurement and the principles of pragmatism.³ First, both pragmatism and performance measurement have a common overall purpose: to challenge conventional wisdom. Many governmental programs have carried over from the past. Well-meaning traditional programs came under attack; the phrase “throwing money at a problem” came into vogue. The feeling was that money was being spent with no discernible positive outcome. A properly developed performance measurement system allows the outcomes of a government program to be known, and subsequently improved or defunded.

A related broad similarity between pragmatism and performance measurement is the notion of putting theory to the practical test. Performance measurement can be viewed as an inquiry into a particular program or intervention. Implicit in performance measurement is the hypothesis that an intervention will have explicit outcomes, such as a reduction in teen pregnancy, or an increase in employment. These hypotheses can then be practically tested in the real world. Doubt about the efficacy of an intervention can be confirmed or denied.

The pragmatic concepts of critical reasoning and the scientific approach are found within the best practices of performance measurement literature (Williams et al., 1994). An intervention should be viewed as an experiment. Care must be taken to ensure, to the extent possible, that the outcomes are a result of the intervention and not other influences.

The literature suggests that the measure development process include an awareness of the full array of measures, consistent with the pragmatic view of the whole. Performance measurement, like the pragmatic approach, should be viewed as a continuing and iterative experimental process. Best practices call for interventions and measures to be assessed annually and revised as needed. Then the whole process begins again with the ability to compare previous data to new data.

Finally, performance measurement works best when the process is open to contributions from others, a major pragmatic requirement. Staff and community can help in the development of a comprehensive and accurate reporting system. In addition, agencies providing similar or related interventions can become part of this “community of inquiry”, as can regulatory and advisory groups.

Constructing Performance Measures: The State of the Art and Science

The same body of literature that delineates best practices in the field also describes serious deficiencies in the process as practiced. The overall success of a performance measurement system hinges, first and foremost, on the quality of performance measures that have been developed (Heinrich, 1999, p.389). Unfortunately, performance measure development remains problematic for a number of reasons, as discussed below. These difficulties have had a significant effect on practicing professionals. Melkers and Willoughby (2001, p.60) found in a survey of state budget workers that “‘difficulties in establishing appropriate performance measures’ and ‘problems in defining performance’ were the most frequently cited problems.”

Why is the science of performance measure development difficult? Government programs have a fundamental problem in measuring success when compared to the business world. In the private sector, there is a single measure of success: profit. No such measure exists for government (Ellwood, 2000, p.320). Government must rely on proxy measures that tend to be more subjective in nature (GAO, 1993, p.2-7).

In some cases, outcomes of public programs simply cannot be measured. Wilson (1989 p.158-171) defines four types of government organizations based on the measurability of outputs and outcomes. In three of the four agency types, outputs and outcomes cannot both be measured.⁴

Another factor that can make performance measurement difficult is the inability to attribute outcomes to, and only to, a specific intervention (GAO, 1993, p.5) Ingraham and Donahue (2000, p.295) cite such outside factors as elections, economic conditions, and the media as having the ability to influence performance.⁵ In

addition, government programs often overlap or duplicate services, making it unclear which intervention is responsible for outcomes (Kravchuk and Schack, 1999, p.354). Also making accountability somewhat more problematic is that agencies that do not provide direct services, but instead contract for same, are one step removed from these services and activities (Joyce, 1993, p.12). Mullen and Magnabosco (1997, p. 4) believe “that no human service official can with fairness be held fully accountable for outcomes, because in the real world, many other factors in addition to the subject program affect outcomes.”

Forging agreement on outcomes can be difficult when many stakeholders are involved. Public policy is inherently political; numerous players need to be considered (Ellwood, 2000, p.330). And these players can have legitimate disagreement. For instance, one stakeholder may think that accessibility is the most important outcome for a bus routing schema. At the same time, another may legitimately believe that efficiency is more important. The failure of the various players to agree on performance measures may actually be a failure to agree on purpose.

Working With Data

In addition to the difficulty in constructing appropriate performance measures, the mechanics of measure development and the use performance data can be problematic. The success of performance measurement is directly related to the ability to manage data, and these systems can consume tremendous amounts of data (GAO, 1993, p.26). Radin (1998) reports that federal agencies have differing abilities to handle data, both in terms of staff skills and necessary hardware.

Williams et al. (1994) caution against the use of data by untrained staff and recommend training in the use and misuse of data. They also recommend that raw data be “filtered” and interpreted by the research department.

Kravchuk and Schack (1996, p.355-356) also fear the misinterpretation of data related to the need to formulate a comprehensive and balanced array of measures. An overly complex and detailed array of measures can cause a program manager to get lost in minutiae. On the other hand, a simple, rudimentary array of measures may not be a valid representation of overall program performance and may direct management focus to achieving these simple objectives at the expense of other important functions that are not represented in the array of measures.

Since this applied research project focuses on a single state’s performance measurement system, it is important to know the state of the art and science in Texas. The following section addresses the setting of this study.

Performance Measurement in Texas

The State of Texas has developed a model integrated performance measure system and takes its place as a national leader (Melkers, 1998, p.67). Since the focus of this study is the Texas performance measurement system, it is important to compare Texas practices and trends with the national literature. Following is a discussion of the local landscape in which this study is conducted.

The State of Texas uses a two-pronged approach to assess agency performance, employing both

performance-based budgeting and performance management (Legislative Budget Board [LBB], 1999b, p.8). In other words, state agencies are held accountable through the linkage between performance and budget; while at the same time, performance data are used to improve agency performance.

In 1991, the Texas legislature passed legislation requiring the linking of strategic plans, including measurable goals and objectives, to the budgeting process. In fact, the strategic plan is the mechanism by which state agencies request funding. Goals, objectives, and performance targets contained in the strategic plan are then included as part of the General Appropriation Act, the means by which agencies are approved for funding. Regular quarterly and annual performance reporting is required (Broom, 1995, p.3-5). The Legislative Budget Board (LBB) has primary responsibility for guiding the development of state agency strategic plans and monitoring achievement of targets associated with performance measures (LBB, 1999a).

The LBB also has certain expectations regarding the process by which measures are developed. First, the responsibility for developing a measurement system falls primarily on state agency management. Management is responsible for determining the key functions that will be measured and for ensuring that measures guide the management of the agency. The LBB must approve measures and revisions, but agency management is responsible for their development (LBB, 1999b, p.7-8).

Secondly, the LBB encourages an inclusive process for measure development. It recommends that the governing board, management, program staff, budget staff, LBB and Governor's Office staff, and agency customers be included in measure development (LBB, 1999b, p.10).

State Auditor's Office Survey

In 1998, the State Auditor's Office (SAO) conducted a survey of Texas state agency administrators regarding their perceptions of the performance measurement system. The results seem to indicate considerable dissatisfaction with the performance measurement process. Selected results of this survey are as follows:

1. Only 58% of the respondents felt performance measures and the strategic plan "always" or "almost always" described what their agencies did.
2. Fewer than half (47%) of the respondents believed that performance measure data was "always" or "almost always" used for agency management.
3. When asked if the key measures represent the most important indicators of agency performance, fewer than half of the respondents (46%) agreed.
4. Only 21% of the respondents said they "always" or "almost always" include performance measures in their subcontracts.
5. When asked to list improvements that could be made to the system, the one most frequently cited was to increase flexibility in adjusting targets, measures, and strategies. The second most frequently was to increase input from state agencies in the development and measurement of targets.

The SAO survey data showed an interesting differential between large and small agencies.⁶ Of the 17 closed-ended questions on the survey, 16 indicated that smaller agencies were less satisfied with the system than were larger agencies. Of the nine questions that pertained directly to performance measurement, small agencies were in each case less approving of the current system when compared to large agencies.

Discussion

Earlier in this presentation, discussion centered around the science-based, best practices of performance measure development. While there is considerable guidance and discussion of the pitfalls of measure development in the literature, this same body of literature indicates that there remain significant problems in the practice of measure development and the use of performance data.

An analogous situation seems to have occurred in Texas. The measure development process as described in LBB documentation embraces many pragmatic principles. However, the SAO survey shows considerable dissatisfaction regarding the process, and the results of the process.

Conceptual Framework and Methodology

Analysis of the tenets of American pragmatism and the best practices of performance measurement brings to light three overarching and related themes: the community of inquiry, the use of scientific methodology, and the on-going use of performance data to inform decision making. In this section, these themes are recast into working hypotheses, which then become the focus of this research. A description and justification of the research methodology follows as well.

Conceptual Framework

The purpose of this applied research project is to explore the extent to which pragmatic practices are employed in the particular area of performance measurement in the State of Texas system.⁷ Working hypotheses are the most common conceptual framework associated with exploratory research (Shields, 1998, p.207) and are utilized in this project.

The first set of hypotheses relates to the presence of community of inquiry principles in the state's performance measure development process. Specifically, the concepts of a mutually responsive atmosphere among the principal players and an open and inclusive process are addressed.

The second set of hypotheses relates to the pragmatic theme of employing scientific principles in performance measure design and definition. Sub-hypotheses center around the use of an ongoing, iterative, and experimental process in measure development, and the quality of measures this process produces. An additional thrust is to determine the abilities of state agency staff to accurately interpret and apply performance measure data.

The third set of hypotheses addresses the use of performance data to attain pragmatic truth. The measures, and the activities and interventions they represent, should be tested and retested in practical application. That is, performance data should be used by program managers to improve operations. After improvements have been made, the ensuing performance data should indicate whether the improvements were successful, and the process begins again. If performance data are not used in this way, performance measurement becomes an exercise with little practical use.

Finally, a fourth, overarching hypothesis pertains to the differences between large and small agencies regarding the presence of pragmatic principles. While the general literature on performance measurement does not seem to address such differences, the Texas SAO survey clearly indicates that small agencies tend to be less

enamored with the performance measurement system than large agencies. For each of the survey items, data will be analyzed separately for large and small agencies. Table 4.1 below details each of the working hypotheses and the literature from which they were drawn.

Methodology

This applied research project addresses the use of pragmatic practices in performance measurement with regard to a particular case: the State of Texas performance measurement system. Since this research explores performance measurement as practiced within state government, it is difficult to separate the phenomenon of performance measurement from the context of government. In such a situation, case study methodology is appropriate (Yin, 1994, p.13).

Yin (1994, p.13) recommends that multiple data sources be used in case studies, and that theoretical propositions guide data collection. In this research, two data collection strategies, surveys and a focus group, are used to explore the working hypotheses discussed above and detailed in Table 4.1.

Surveys are considered the most efficient and reliable way to garner standardized information from large numbers of respondents. However, validity can be a concern because respondents are given only a few response options. These options may not convey the exact views of the respondent, nor are they likely to capture all that the respondent thinks about a given question (Bobbie, 1998). For this reason, a focus group is also employed.

Focus groups are often used in combination with other research methods. The interaction among members of a focus group can add a depth of understanding unavailable using other techniques. The use of focus groups in conjunction with data garnered from other methods can provide a “more accessible understanding” when findings are presented (Morgan, 1997, p30).

Survey

Survey items were developed around the first three hypotheses shown in Table 4.1 above. Each hypothesis comprises two to four sub-hypotheses, and each survey item is attached to a specific sub-hypothesis. The fourth hypothesis, regarding the differences between small and large agencies, pertains to all the survey items. The survey⁸ was developed by the author and was pretested by two former state agency employees.

The intended unit of analysis for the survey data is the state agency. Owing to the anonymous nature of the survey, it is possible, however, that surveys were copied and provided to other staff members, resulting in multiple responses from a single agency. Collaboration between staff members on a single agency response is also not precluded, although this situation is not considered a problem for interpreting results. Future research along these lines, however, could benefit from a stronger emphasis on guaranteeing a single response from each agency. Survey responses are held in strict confidence.

A five-response-option Likert scale is employed for each of the 19 survey items. Two additional items on the survey are descriptive: agency size and agency type. The coding for each survey item is as follows: Strongly Agree = 5, Agree = 4, Neutral = 3, Disagree = 2, and Strongly Disagree = 1. Survey data were entered into an SPSS database. Ambiguous responses were not recorded; the item in question was left blank. However, the rest of the valid responses were entered for that observation.

Table 4.1 Working Hypotheses

<u>The Performance Measure Development Process</u>			
WH1: The Texas state agency performance measure process employs “community of inquiry” principles.			
	Literature Sources	Subhypothesis	Research Methods
WH1a	Shields (forthcoming) Ellwood (2000)	The measure development process involves active input from a broad array of sources.	Survey Q1 & 2 Focus Group
WH1b	Shields (forthcoming) Evans (2000)	Participants are mutually responsive.	Survey Q3-Q6 Focus Group
<u>The Use of Scientific Principles</u>			
WH2: Sound scientific practices are employed in performance measure use and development.			
	Literature Sources	Subhypothesis	Research Methods
WH2a	Shields (1998) Williams (1994)	The measure development process is ongoing, experimental, and iterative.	Survey Q7-Q9 Focus Group
WH2b	SAO (1998) GAO (1993) Ingraham (2000) Kravchuk (1999) Heinrich (1999)	Performance measures are an accurate representation of agency function.	Survey Q10 Focus Group
WH2c	SAO (1998) GAO (1993) Kravchuk (1996)	The arrays of measures per agency accurately reflect the overall mission.	Survey Q11-Q13 Focus Group
WH2d	Williams (1994)	State agencies have sufficient expertise to interpret and use performance measure data.	Survey Q14-Q16 Focus Group
<u>The Use of Performance Measurement Data</u>			
WH3: Performance data are used to guide agency operations.			
	Literature Sources	Subhypothesis	Research Method
WH3a	Jordan (1999)	Performance data are used to guide allocation of funds.	Survey Q17 Focus Group
WH3b	Jordan (1999)	Performance data are used to improve agency operations.	Survey Q18 Focus Group
WH3c	SAO (1998) L. Martin (1997)	Performance data are used to improve contractor performance.	Survey Q19 Focus Group
<u>Large versus small agencies</u>			
WH4: Pragmatic principles are less evident in small state agencies compared with large agencies.			
	Literature Sources	Subhypothesis	Research Methods
WH4a	SAO (1998)	Community of inquiry principles are less evident in small state agencies compared to large agencies	All Survey Questions
WH4b	SAO (1998)	Scientific principles are less evident in small state agencies compared with large agencies.	All Survey Questions
WH4c	SAO (1998)	Use of performance data to guide agency operations is less evident in small state agencies compared with large agencies.	All Survey Questions

Focus Group

A focus group was convened to assist in interpreting survey data. Focus group invitations were extended to 25 local email respondents. Agencies that mailed in surveys were largely anonymous, and were therefore not offered the opportunity to participate in the focus group.⁹ Only four of the invited respondents agreed to participate, one of whom did not attend because of illness. The attendees were all managers intimately involved with performance measurement at their agency. Each represented a large state agency.

Focus group participants were emailed a prospectus of the study and available data prior to the meeting. At the start of the focus group, the participants were introduced to each of the global hypotheses and sub-hypotheses. The data were reviewed in order and participants were asked to comment and offer any additional opinions and observations regarding each sub-hypothesis. They were also asked to offer any suggestions for improving the Texas performance measurement system. A \$25 stipend was offered for participation; all participants declined. The session was audio taped.

Sample

The study population for this research includes all state agencies listed on the State of Texas webpage¹⁰ as of September 23, 2001. Quasi-state agencies¹¹ which appeared on the webpage were excluded, as was the Legislative Budget Board for obvious reasons. Surveys were sent to all other listees. Three surveys were discarded, and excluded from study calculations, because the survey was inappropriate or misunderstood.¹²

The official agency email address as listed on the State of Texas webpage was used to transmit the survey. In cases where email addresses were not available, surveys were mailed to the executive directors of the agencies. Ninety-seven (97) surveys were mailed; one-hundred-seventeen (117) were emailed for a total of 214. Eighty-six (86) acceptable surveys were returned, yielding a response rate of 40%.

Analysis of Sample

A chi-square analysis¹³ was conducted to determine if the respondents were representative of the sample. Since the unit of analysis in this study is the state agency, agency type was used as the dependent variable. Analysis indicated that the null hypothesis (agency type distribution for the respondents would not be different from the sample) could be rejected. For this reason, any inferences regarding the State of Texas performance measurement system as a whole should be made only with extreme caution.

Survey respondents were asked to categorize their agencies as large or small based on whether their total budget exceeded \$40 million or not. Just over half (52.6%) that answered this question were large agencies.

Results

Survey and focus group data are presented below, organized by working hypothesis and sub-hypothesis. Summary data for each of the 19 survey items are presented in tabular form. Survey results are used to determine the level of support for each of the sub-hypotheses. Focus group data are included in the narrative discussion of each survey item.

The fourth working hypothesis (large versus small agencies) concerns all survey items. Hence, the data associated with this hypothesis are presented and discussed along with each of survey questions. That is to say, there is no separate section for reporting results regarding the fourth working hypothesis.

Survey Recoding

A survey question is considered to support the associated working hypothesis if at least 50% of the respondents agree or strongly agree with the item. In other words, a respondent has to affirmatively state some level of agreement to be counted in support. For this reason the original responses have been recoded. “Strongly agree” and “Agree” responses have been summed to create a “Combined Agree” value. Similarly,

“Strongly Disagree” and “Disagree” values have been merged and counted as “Combined Disagree”. Neutral responses remain the same.

When this combined agree score is 50% or more, the percentage is annotated with an asterisk (*) to denote support for the sub-hypothesis according to the protocol of this study. Similarly, when small agency percentages are less than those of large agencies, a double asterisk (**) is affixed to indicate support for the fourth working hypothesis (WH4). Certain explanatory survey items provide no evidence, one way or the other, regarding the hypothesis, but serve to add color to the findings. In these cases, as noted in the findings narrative, the 50% rule will not apply.

A mean value of the responses to each question is also presented in the tables. The mean values, unlike the percentages, are based on the original survey coding to provide a more accurate representation. For a given survey item, if the respondents were equally divided among the five response options, the mean would equal three; this value is used as a benchmark when considering mean scores. Many researchers do not recommend taking the mean of ordinal data (Gravetter and Wallnau, 1999, p.71-72). In this research, however, the mean will be only used for explanatory purposes in association with the primary analysis method described above.

Working Hypothesis 1: Performance Measure Development

This first working hypothesis presupposes that community of inquiry principles are present in the State of Texas performance measurement development process. In other words, the community comes together to address the problem of measure definition. The first sub-hypothesis concerns the input, both from outside the agency and from within, that goes into measure development. The second sub-hypothesis addresses the mutual responsiveness in the relationship between the principal parties in the process, namely, the state agencies and the LBB.

Working Hypothesis 1a: Input from Stakeholders

The premise that measure development involves broad input is only partially supported by the data. Responding agencies, on the whole, do not reach outside their confines to solicit input, but do significantly involve agency staff in the process. As indicated in Table 5.1 below, only one-third of the respondents agreed or strongly agreed that their agency went to outside parties for assistance in performance measure development, while just over 38% disagreed or strongly disagreed. The mean value came in under the benchmark at 2.9. Large agencies agreed with this proposition a striking 17 percentage points higher than small agencies. One focus group member stated the opinion that in some cases outside parties had nothing to offer the performance measurement development process. It was acknowledged that this may differ from agency to agency. Another member felt that outside participation was only valuable in a global sense. That is, useful input could be gained from the outside regarding overarching policy direction, but the crafting of performance measures required skills that outsiders were not likely to have.

The performance measurement development process did seem to include line staff from within the agency. Just over 68% of the respondents supported the premise that agency staff are involved in the process. The mean is considerably above the benchmark at 3.5. The differential between large and small agencies was less pronounced at 11.2%.

Table 5.1 Input from Stakeholders

WH1a	The measure development process involves active input from a broad array of sources.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
	Input from outside stakeholders n = 84	33.3	2.89	40.9	23.7**
	Input from agency line staff n = 85	68.2*	3.54	72.7	61.5**

* findings support the working hypothesis

** findings support WH4

Working Hypothesis 1b: The Relationship between the Principle Players

The second sub-hypothesis pertains to the relationship between state agencies and the LBB. When asked if staff of the respective parties were equal in the performance measure process, over 56% combined to disagree, while only 19.5% combined to agree. Only 13.5% of the small agencies combined to agree with this premise. This measure establishes that the respondents, on the whole, do not believe the process was equally shared, but does not indicate in which direction they felt the balance tipped. The next three survey items help explain this aspect.

As noted above, LBB guidelines place the burden of measure development on the state agencies. These last three survey items address actual practices in this regard. If the relationship is mutually responsive, one would expect LBB to acquiesce to agency requests regarding the initiation, modification, and elimination of measures. The data indicate, LBB guidelines notwithstanding, that only 38% of the respondents agreed that their state agencies initiate measure development. Similarly, there was not majority support for the premise that the LBB agrees to modify or eliminate performance measures at state agency request; combined agree percentages were 42.7% and 32.5%, respectively. Small agencies agreed with these premises at a rate of about 14 percentage points lower.

Table 5.2 The Relationship between the Players

WH1b	Participants are mutually responsive.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
	LBB and agency staff are equals. n = 82	19.5	2.48	20.9	13.5**
	Measures are initiated by agency staff. n = 84	38.1	2.87	37.8	35.1**
	LBB agrees to modify measure definitions. n = 82	42.7	3.07	48.8	35.1**
	LBB agrees to eliminate measures. n = 82	32.5	2.92	38.6	24.3**

* findings support the working hypothesis

** findings support WH4

The focus group, however, had a more generous view of the LBB's relationship with state agencies. One

veteran state agency staffer felt that the relationship between the two principal players has improved over the last five or six years; he felt that LBB might be suffering from lingering ill feelings from times past. He also said that the relationship between an agency and the LBB was colored by the approach of the assigned LBB staff. Some staff, in his view, were more egalitarian than others.

The focus group agreed that LBB staff were much more likely to modify or eliminate measures at the beginning of the budget cycle when appropriations, measures, and performance targets are set for the biennium. Once the biennium was underway, LBB staff were less inclined to accept any change or elimination of measures, although focus group members did report success at changing targets midstream.

In sum, then, the survey data provide only weak support for the existence of a mutually responsive relationship between the LBB and state agencies. This finding is tempered somewhat by the views of the focus group.

Working Hypothesis 2: The Use of Scientific Principles

This working hypothesis comprises four sub-hypotheses, all of which combine to provide an assessment of the use of scientific principles in the performance measurement process. Specifically, the four sub-hypotheses address the experimental nature of performance measurement as practiced, the accuracy of measures, both individually and as a whole, and agency staff’s ability to use performance data.

Working Hypothesis 2a: The Experimental Process

The first sub-hypothesis regards the cyclical development of performance measures. If an ongoing, experimental, iterative, approach is used in performance measurement, one would expect measures to change as agency functions change. One would also expect the application of performance data to modify measure definitions and targets. As indicated in Table 5.4 below, each of the premises is supported by at least 50% of the respondents.

Exactly half of the respondents felt that agency measures change as agency functions change, with a mean score of 3.4. In addition, over 56% of the respondents reported that performance data had been used to modify measure definitions, while a robust 76.5% reported that performance targets had been changed in that period. The corresponding mean values of 3.4 and 3.9 are congruent with these data.

Table 5.4 Use of scientific practices

WH2a	Sound scientific practices are employed in performance measure use and development.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
	Performance measures change as agency functions change. n = 82	50.0*	3.38	54.8	42.1**
	Data were used to modify definitions in last 2 years. n = 83	56.6*	3.40	61.9	53.8**
	Data were used to modify targets in last 2 years. n = 85	76.5*	3.89	82.2	71.8**

* findings support the working hypothesis

** findings support WH4

Smaller agencies are again less supportive of the sub-hypothesis. Each question associated with this sub-hypothesis has lower combined agree scores for smaller agencies.

Working Hypothesis 2b: Accuracy of measure definitions.

Having established that scientific principles are, according to the respondents, largely in place in the performance measurement development process, this sub-hypothesis addresses one of the end results of the process: the accuracy of the measures. Just over one-half of the respondents combined to agree that performance measures are accurate representations of agency functions. A mean score of nearly 3.3 supports this premise as well. Small agencies, however, are not supportive of this proposition, scoring 4.6 percentage points below large agencies, and dropping below the 50% benchmark. See Table 5.5 below for details.

Table 5.5 Accuracy of performance measures.

WH2b	Performance measures are an accurate reflection of agency functions.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
	Measures accurately represent agency functions. n = 85	50.6*	3.27	53.3	48.7**

* findings support the working hypothesis

** findings support WH4

Noting that a scant majority supported this premise, one of the focus group members offered the opinion that performance measures did a much better job of measuring outputs than outcomes.

Working Hypothesis 2c: Accuracy of combined measures

While individual measures are considered accurate by a bare majority of the respondents, they are considerably less positive when considering all agency performance measures as a whole. Only about one-third of the respondents (32.9%) combined to agree that the full of array of measures accurately represents overall agency functioning. Large agency support for this proposition is almost 10 percentage points higher than small agencies. The lack of support for this sub-hypothesis does not seem related to an overabundance or scarcity of measures as indicated in the two explanatory items¹⁴ that follow. Just over 7% of the respondents felt there were too few measures, while nearly 39% thought there were too many. Large agencies were much more inclined to feel that they had too many measures, scoring 18.5 percentage points higher than small agencies. See Table 5.6 below for detailed data.

Table 5.6 Accuracy of Combined Measures

WH2c	The arrays of measures per agency accurately reflect the overall mission.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
	Measures accurately represent overall agency mission n = 85	32.9	2.89	37.8	28.2**
Explanatory	My agency has too many measures. n = 85	38.8	3.19	46.7	28.2
Explanatory	My agency has too few measures. n = 85	7.1	2.28	6.7	7.7

* findings support the working hypothesis ** findings support WH4

The focus group and a survey respondent reported that their agencies have an alternate performance measurement system in place. The focus group member felt the LBB measures were incomplete, and additional data was needed by agency management. This phenomenon often takes place at the department or division level, rather than agency-wide.

Finally, the focus group felt that a comprehensive array of measures would offer agency staff clear guidance on their overall mission and what are considered the most important agency functions.

Working Hypothesis 2d: Agency staff expertise in use of data

In order for scientific principles to be successfully applied, agency staff need the ability to work with and understand performance measures and performance data. The questions associated with this working hypothesis address the abilities of state agency staff in this regard, as seen through the respondents' eyes. The respondents overwhelmingly reported high skill levels. Over three-fourths (76.5%) combined to agree that agency staff had the necessary skills to develop quality measures. The respondents combined to agree at very high levels that state agency staff had the requisite abilities to interpret and apply performance data; the combined agree percentages were 89.4% (mean = 4.1) and 85.7% (mean = 4.0), respectively.

When small agencies are differentiated from large agencies, the data still indicate widespread support for this sub-hypothesis. It is noted, however, that small agencies support the notion that agency staff have the expertise to develop quality measures at about 10 percentage points lower than large agencies. See Table 5.7 below.

Table 5.7 Agency staff expertise in use of data

WH2d	State agencies have sufficient expertise to interpret and use performance measure data.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
	Staff have the skill to develop quality measures n = 85	76.5*	3.86	82.2	71.8**
	Staff have the ability to interpret performance data n = 85	89.4*	4.11	93.3	87.2**
	Staff have the ability to apply performance data n = 84	85.7*	4.00	86.7	86.8

* findings support the working hypothesis ** findings support WH4

Working Hypothesis 3: The Use of Performance Measure Data

The use of performance measure data to modify measures and targets has been addressed above when discussing the hypotheses regarding scientific processes in performance measurement. This third hypothesis is concerned with the use of data in operational areas, namely, allocation of funds, agency operations, and contractor performance. Each of the sub-hypotheses has a single associated question.

Working Hypothesis 3a: Use of performance data to guide allocation.

An even 50% of the respondents combine to agree that their agencies use performance measure data in allocation of funds. The mean value of 3.3 for this variable is slightly higher than the benchmark value. Small agencies supported this proposition at 52.6%, nearly 11% higher than large agencies. Overall, this hypothesis is supported, if barely. It should be noted that four observations were eliminated from these calculations because of a “Not Applicable” response. See Data Table 5.8 below for more detail.¹⁵

Working Hypothesis 3b: Use of performance data to improve operations.

The respondents reported considerable use of performance data to improve agency services, as nearly 59% combined to support this hypothesis. The mean value is congruent at almost 3.4. Small agencies agreed with this premise at a slightly higher rate than large agencies.¹⁶ Detailed data are found in Table 5.8.

Working Hypothesis 3c: Use of performance data to improve contractor performance.

The data associated with this sub-hypothesis are strikingly different from the previous two. Only 26.6% of the respondents combined to support the hypothesis that state agencies use performance data to improve contractor performance. A nearly equal amount (28.3%) combined to disagree. However, the veracity of the data is questionable. There were seven missing observations and 19 “Not Applicable” responses. This item had the highest number of neutral answers, which may mean that respondents indicated “Neutral” when their agency did not outsource any services, instead of the appropriate “N/A”. Nearly equal agreement and disagreement percentages with a high number of neutral responses (value = 3) pull the mean (3.0) toward the benchmark.

Small agency agreement with this hypothesis was very low, almost 27 percentage points below the large agency rate. See Table 5.8, below.

Table 5.8 Use of performance data to affect agency operations.

WH3	Performance data are put to optimal use by state agency staff.	All agencies		Large Agencies	Small Agencies
		% Agree	Mean	% Agree	% Agree
WH3a	Data are used to guide allocation of funds n = 78	50.0*	3.32	41.9	52.6
WH3b	Data are used to improve agency operations. n = 85	58.8*	3.39	57.8	59.0
WH3c	Data are used to improve contractor performance. n = 60	26.7	3.00	32.6	5.7**

* findings support the working hypothesis

** findings support WH4

Conclusions and Recommendations

The purpose of this applied research project is to determine the extent to which the principles and practices of American pragmatism can be found in the art and science of performance measurement, as practiced in the State of Texas. A review of the literature has shown striking similarities between basic pragmatic tenets and the best practices of performance measurement. At this level, it seems clear that pragmatic principles are indeed implied, if not acknowledged, in the area of performance measurement. It also seems clear from the present research that there is a significant presence of basic pragmatic principles in the everyday practice of performance measurement in Texas.

To consistently evaluate support for the working hypotheses, a coding and rating schema was devised. Support ratings for the sub-hypotheses were determined by the support of the associated survey items (See Table 6.1 below). Numeric values were assigned to the different ratings. These assigned values were then averaged to reckon the level of support for each global hypothesis. When the associated sub-hypotheses were inconsistent in their support for a hypothesis, the support rating was considered mixed (e.g. mixed majority support).

Table 6.1 Sub-hypothesis Support Rating

No support – none of the survey items support sub-hypothesis.
Weak support – at least one, but fewer than half of the survey items support sub-hypothesis.
Majority Support – half or more, but not all survey items support sub-hypothesis.
Complete Support – all of the survey items support sub-hypothesis.

For the fourth hypothesis (WH4), this method is taken a step further. WH4 concerns the difference in percentage scores between small and large agencies for each survey question. Since this hypothesis embodies the other three, the average values associated with each sub-hypothesis are averaged to determine the overall level of support for WH4.

Conclusions

Each of the working hypotheses is addressed in turn. Results are shown in Tables 6.2 and 6.3 below.

Working Hypothesis 1: Performance Measure Development

The Community of Inquiry concept received only mixed weak support from the respondents. The sub-hypothesis regarding input into the performance measure process received majority support. The respondents reported that agency line staff were involved in the process, but the proposition regarding outside input did not receive significant support. Further, according to the respondents, the relationship between the agencies and LBB seems to be tilted to one side, with LBB enjoying more influence.

Working Hypothesis 2: The Use of Scientific Principles

There is significant support for the propositions regarding the use of scientific practices in the performance measurement development process. There seems to be a cyclical retooling of both targets and measures as performance data is fed back into the system. It also appears, in the most robust findings in this study, that agency staff have considerable skill in developing measures, interpreting data, and applying findings. While the

findings show significant scientific application, the respondents, on the whole, do not believe that the body of measures for their agency represents true overall agency functioning. All told, this second working hypothesis received mixed majority support.

Working Hypothesis 3: The Use of Performance Measure Data

The majority of evidence from the respondents supports the proposition that performance data are used to modify agency functioning on a regular basis. Performance data were thought to guide allocation of funds and improvement of services, but not contractor performance. This is a significant finding because it sets up a cyclical approach whereby data are gathered, analyzed, and applied, positioning the agency to measure the effects of their modifications when the process is repeated.

Working Hypothesis 4: Large versus Small State Agencies

Finally, there is considerable support for the supposition that pragmatic principles would be less evident in small state agencies when compared with large ones. Regarding community of inquiry principles, every survey item attached to both sub-hypotheses indicated that small agencies supported the hypothesis less than large agencies. The small versus large agency sub-hypothesis was overwhelmingly supported.

Three of the four sub-hypotheses regarding the use of scientific principles were completely supported. The remaining sub-hypothesis, concerning staff expertise in using performance data, received majority support. Overall, the data regarding the use of scientific principles offered mixed majority support to the small versus large hypothesis.

Finally, the sub-hypothesis concerned with the use of data to affect agency operations received only mixed weak support. Two of the sub-hypotheses offered no support for the overall hypothesis. The sub-hypothesis concerning contractor performance was completely supported.

When the support for the three sub-hypotheses is considered together, it translates to mixed majority support for Working Hypothesis 4 (WH4). See Table 6.3 for detailed data.

Table 6.2 Summary of Findings – WH1, WH2, and WH3

Working Hypotheses		Sub-hypotheses	Hypotheses
WH1	The Texas state agency performance measure process employs “community of inquiry” principles.		mixed weak support
	WH1a	Broad input in performance measure development	majority support
	WH1b	Mutually responsive relationship between participants	no support
WH2	Sound scientific principles are used in performance measure use and development		mixed majority support
	WH2a	The process is ongoing, experimental and iterative	complete support
	WH2b	Measures are an accurate representation of agency function.	complete support
	WH2c	Arrays of measures accurately represent the overall mission	no support
	WH2d	Agencies have sufficient expertise in use of performance data.	complete support
WH3	Performance data are put to optimal use by state agency staff.		mixed majority support
	WH3a	Performance data are used to guide allocation of funds.	complete support
	WH3b	Performance data is used to improve agency services.	complete support
	WH3c	Performance are used to improve contractor performance.	no support

Table 6.3 Summary of Findings – WH4

Working Hypotheses		Sub-hypotheses Components	Sub-hypotheses	Hypothesis
WH4	Pragmatic principles are less evident in small state agencies compared with large agencies			mixed majority support
WH4a	Community of inquiry principles are less evident in small state agencies compared to large agencies.		complete support	
WH4a(1)	Broad input in performance measure development	complete support		
WH4a(2)	Mutually responsive relationship between participants	complete support		
WH4b	Scientific principles are less evident in small state agencies compared with large agencies		mixed majority support	
WH4b(1)	The process is ongoing, experimental and iterative	complete support		
WH4b(2)	Measures are an accurate representation of agency function.	complete support		
WH4b(3)	Arrays of measures accurately represent the overall mission	complete support		
WH4b(4)	Agencies have sufficient expertise in use of performance data.	majority support		
WH4c	Use of performance data to guide agency operations is less evident in small state agencies compared with large agencies.		mixed weak support	
WH4c(1)	Performance data are used to guide allocation of funds.	no support		
WH4c(2)	Performance data are used to improve agency services.	no support		
WH4c(3)	Performance data are used to improve contractor performance.	complete support		

Discussion

When considered as a whole, the preponderance of evidence supports the idea that selected pragmatic principles are present in the State of Texas performance measurement system. The data also indicate that this is less so for small agencies when compared to large agencies.

It is noted, however, that several of the survey items were supported by the barest of margins. The 50% criterion was met exactly in two cases and barely met in another. While these items did support the hypotheses under the decision rule established in this research, it should be recognized that there was a sizable minority. That is to say, that even though a hypothesis was supported according to the protocol in this research, there appears to be room for considerable improvement, at least in the eyes of some.

Recommendations

The present research indicates that there are two apparent areas where the greater use of pragmatic principles would improve the functioning of the State of Texas performance measurement system.

1. The survey data strongly suggest, and the focus group concurs, that the system would be improved, at

least from the state agency perspective, if LBB staff would take a more flexible and egalitarian approach to the performance measurement process. By embracing the community of inquiry approach, a better product, with less resentment, could be achieved

2. The principal parties should pay more attention to the full array of measures for a given agency. The data would seem to suggest that there are important areas of functioning that are not measured. The performance measurement system would likely have more credibility with state agencies if measures accurately reflected overall functioning. In addition, staff could look to the body of measures for overall guidance and understanding of agency mission.

Recommendations for Further Study

1. Pragmatic principles were much less evident in small agencies regarding community of inquiry principles and the use of a scientific approach. Study of the reasons for these differences might prove rewarding.

2. The data showed that outside stakeholders were not, by and large, included in the performance measure development process. The focus group, however, thought that these stakeholders could contribute only with global policy advice, if at all. The usefulness of these stakeholders in the performance measurement process could be a subject of further inquiry.

3. Agencies reported that when considered as a whole, measures did not reflect overall agency functioning. It is possible that at least part of the reason is because some activities and functions are inherently difficult to measure. An inquiry into this aspect of performance measurement might enrich the field.

Endnotes

¹ Shields, forthcoming, p.77; Snider, 2000a, p.336; Zanetti and Carr, 2000, p.436

² Shields, 1996; Stever, 2000; and Svava, 2001.

³ Following is a summary of performance measurement best practices found in the literature, and their link to pragmatism. Information regarding performance measurement is drawn primarily from the following sources: GAO (1993), Kravchuk and Schack (1996), Osborne and Gaebler (1992), Williams, McShane, and Sechrist (1994).

⁴ For instance, Wilson defines procedural organizations as able to measure outputs but not outcomes. A drug and alcohol abuse prevention program for young children is a good example of such an organization. The hoped-for-result might be that kids will not use drugs and alcohol while underage, and will use alcohol responsibly when old enough. The success of the program will not be apparent for some years. It is easy, however, to count outputs such as program enrollment, attendance, and completion.

⁵ Note that these factors are the same as threats to internal validity in experimental design. That is, an alternate explanation cannot be ruled out.

⁶ A large agency's total budget is greater than or equal to \$40 million; a small agency is less than \$40 million.

⁷ The present research differs from the State Auditor's Office survey in the emphasis placed on detecting implied pragmatism within the performance measurement system. The SAO survey was concerned primarily with satisfaction of state agencies with the strategic planning and performance measurement systems. There are common areas of inquiry, particularly regarding the accuracy of measures. However this research emphasizes community of inquiry principles, and actual use of performance data (behavior), instead of satisfaction.

⁸ The survey is located in Appendix A

⁹ Agencies that mailed back the survey typically had no identifying information on the survey. Agencies that returned the survey via email could be identified by the email address. Email respondents whose agency was headquartered in Austin were invited, via return email, to the focus group.

¹⁰ <http://www.tsl.state.tx.us/trail/agencies.html>

¹¹ Lower Colorado River Authority, Texas Guaranteed Student Loan

¹² One agency was part of the executive branch and had no measures, another was no longer in existence. A third survey was discarded because the respondent indicated (incorrectly) that his organization was not a state agency and, thus, had no measures.

¹³ Chi-square analysis can be found in Appendix B.

¹⁴ Note that the 50% decision rule does not apply to explanatory items.

¹⁵ Survey items 17 and 19 may have been confusing for the respondents. In a case where an agency does not allocate funds or employ subcontractors, the appropriate response is 'N/A'. However, in retrospect, it is easy to imagine a response of 'neutral' or 'strongly disagree' in such cases. Data for WH3a and WH3c should be interpreted with caution

¹⁶ It is interesting to note that while smaller agencies indicated a lower level of support for the validity of performance measures, they actually report using the data at a slightly higher rate. The author speculates that a possible reason for this is that smaller agencies are "less bureaucratic" than larger agencies. It could be that data is more easily shared in smaller agencies.

Bibliography

- Allen, J. R. (1996). The uses of performance measurement in government. Government Finance Review, 12(4), 11-15.
- Babbie, E. (1998). The Practice of Social Research. (8th ed.) Belmont, CA: Wadsworth Publishing Company.
- Broom, C. A. (1995, Winter). Performance-based government models: building a track record, Public Budgeting & Finance, 3-17.
- Caro, R. A. (1994), The Power Broker. New York: Alfred A. Knopf
- Current, R. N., Williams, H. T., Freidel, F., & Brinkley, A. (1987). American History: A Survey. (7th ed., Vol II). New York: Random House.
- Dewey, J. (1939). Creative democracy – the task before us. In Hickman, L. & Alexander, T. (Eds.), The Essential Dewey: Volume 1: Pragmatism, Education, Democracy (pp.340-344).
- Ellwood, J. W. (2000). Prospects for the study of the governance of public organizations. In Heinrich, C. J. & Lynn, L. E. (Eds.), Governance and Performance (pp.292-318).
- Evans, K. G. (2000). Reclaiming John Dewey: Democracy, inquiry, pragmatism, and public management, Administration & Society, 32, 308-328.
- Forrester, J. P., & Adams, G. B. (1997). Budget reform through organizational learning: Toward an organizational theory of budgeting, Administration & Society, 28, 466-488.
- Garrison, J. (2000). Pragmatism and public administration, Administration & Society, 32, 458-477.
- Gravetter, F. & Wallnau, L. (1999) Essential of Statistics for the Behavioral Sciences. (3rd ed) Brooks/Cole: Pacific Grove, CA.
- Government Accounting Office (1993). Using performance measures in the federal budget process.
- Heinrich, C. J. (1999). Do government bureaucrats make effective use of performance management information?, Journal of Public Administration and Theory, 9, 363-393.
- Ingraham, P. W. & Donahue, A. K. (2000). Dissecting the black box revisited: Characterizing government management capacity. In Heinrich, C. J. & Lynn, L. E. (Eds.), Governance and Performance (pp.319-335).
- Jordan, M. M. & Hackbart, M. M. (1999, Spring). Performance budgeting and performance funding in the states: A status assessment, Public Budgeting & Finance, 68-88.
- Joyce, P. G. (1993, Summer). Using performance measures for federal budgeting: Proposals and prospects, Public Budgeting and Finance, 3-17.
- Kravchuk, R. S. & Shack, R. W. (1996). Designing effective performance-measurement systems under the government performance and results act of 1993, Public Administration Review, 56, 348-358.
- Legislative Budget Board (1999a). Budget 101: A Guide to the Budget Process in Texas. [On-line] Available: <http://www.lbb.state.tx.us/>

Legislative Budget Board (1999b). Guide to Performance Measure Management [On-line] Available: <http://www.lbb.state.tx.us/>

Legislative Budget Board (2000). Preparing and Submitting Agency Strategic Plans: Fiscal Years 2001-05. [On-line] Available: <http://www.lbb.state.tx.us/>

Martin, L. L. & Kettner, P. M. (1997). Performance measurement: The new accountability, Administration in Social Work, 21, 17-29.

Martin, M. (1978). The Dictionary of American History. New York: Dorsett Press.

Melkers, J. E., & Willoughby, K. G. (1998). The state of the states: Performance-based budgeting requirements in 47 out of 50, Public Administration Review, 58, 66-73.

Melkers, J. E., & Willoughby, K. G. (2001). Budgeters' views of state performance-budgeting systems: Distinctions across branches, Public Administration Review, 61, 54-64.

Morgan, D. L. (1997) Focus Groups as Qualitative Research. (2nd ed.) Thousand Oaks, CA: Sage Publications.

Mullen, E. J. & Magnabosco, J. L. (1997) Outcomes Measurement in the Human Services. Washington, D. C.: National Association of Social Workers.

Osborne, D., & Gaebler, T. (1992). Reinventing Government: How the entrepreneurial spirit is transforming the public sector from schoolhouse to statehouse, city hall to the pentagon. Reading, MA: Addison-Wesley Publishing Company, Inc.

Radin, B. A. (1998). The Government Performance and Results Act (GPRA): Hydra-headed monster or flexible management tool?, Public Administration Review, 58, 307-316.

Rossi, P. H. (1997). Program outcomes: Conceptual and measurement issues. In Mullen, E. J. & Magnabosco, J. L. (Eds.), Outcomes Measurement in the Human Services. Washington, D. C.: National Association of Social Workers.

Sahakian, W. S. (1968). History of Philosophy. New York: Barnes & Noble Books.

Shields, P. M. (1996). Pragmatism: Exploring public administration's policy imprint, Administration & Society, 28, 390-41.

Shields, P. M. (1998). Pragmatism as philosophy of science: A tool for public administration. In White, J. D. (Ed.), Research in Public Administration, 199-230.

Shields, P. M. (forthcoming). The community of inquiry: Classical American pragmatism and public administration. Administration & Society.

Snider, K. F. (2000a). Expertise or experimenting?: Pragmatism and American public administration, 1920-1950, Administration & Society, 32, 329-354.

Snider, K. F. (2000b). Response to Stever and Garrison, Administration & Society, 32, 487-489.

Stever, J. A. (2000). The parallel universes: Pragmatism and public administration, Administration & Society, 32, 453-457.

State Auditor's Office (1998). Performance-Based Budgeting Survey Results. [On-line] Available:
<http://www.lbb.state.tx.us/>

Svara, J. H. (2001). The myth of the dichotomy: Complementarity of politics and administration in the past and future of public administration, Public Administration Review, 61, 176-182.

Waldo, D. (1984). The Administrative State. (2nd ed.) New York: Holmes and Meier.

Williams, F. P., McShane, M. D., & Sechrest, D. (1994). Barriers to effective performance review: The seduction of raw data, Public Administration Review, 54, 537-542.

Wilson, J. Q. (1989). Bureaucracy: What government agencies do and why they do it. Basic Books.

Yin, R. Q. (1994). Case Study Research: Design and Methods. Thousand Oaks, CA: Sage Publications.

Zanetti, L. A. & Carr, A. (2000). Contemporary pragmatism in public administration, Administration & Society, 32, 433-452.

Appendix A

Performance Measure Survey

1	My agency solicits input from interested outside parties in developing performance measures. (e.g. contractors, community)	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
2	Agency line staff are involved in performance measure development.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
3	Agency staff and LBB staff are equals in performance measure development.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
4	Most performance measures are initiated by agency staff rather than LBB staff.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
5	LBB staff agree to agency requests to modify performance measure definitions.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
6	LBB staff agree to agency requests to eliminate performance measures as needed.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
7	Performance measures change as agency functions change.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
8	Performance data were used to modify measure definitions in the last two years.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
9	Performance data were used to modify measure targets in the last two years.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
10	Performance measures accurately represent important agency functions.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11	When all performance measures are considered together, they accurately represent overall agency functioning.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
12	My agency has too many measures.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
13	My agency has too few measures.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
14	Staff at my agency have the necessary skills to develop quality performance measures.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
15	Staff at my agency have the ability to interpret performance measure data.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
16	Staff at my agency have the ability to apply performance measure data.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
17	Performance data has been used for allocation of funds during the last two years.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
18	Performance data has been used to improve agency services during the last two years.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
19	Performance data has been used to improve contractor performance in the past two years.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
20	My agency's total budget is greater than \$40 million.	Yes	No	Don't Know			
21	Circle whichever category best describes your agency's function.	General Government Education Medical Institution Public Safety Business and Econ. Dev.		Health and Human Services University Judiciary Natural Resources Regulatory			

Appendix B

Chi-square Analysis

Analysis indicated that four of the 10 cells had expected frequencies less than five, invalidating the chi-square procedure. The data were then recoded to merge these four cells into one. The subsequent chi-square analysis indicated that the null hypothesis (agency type distribution for the respondents would not be different from the sample) could be rejected, $\chi^2(6, n = 81) = 30.08, p. < .00005$.

Table 4.2 Chi-Square Analysis of Respondents

Agency type (recoded)	Cases		Residual
	Observed	Expected	
General Government	6	7.95	-1.95
Education	17	6.43	10.57
Health & Human Svcs.	13	6.81	6.19
University	10	18.93	-8.93
Judiciary	4	7.57	-3.57
Regulatory	17	15.9	1.1
Med, Pub Safety, Bus & Econ Dev and Natural Resources	14	17.41	-3.41
Total	81		
	Chi-Square	D.F.	Significance
	30.0818	6	.0000